Fitting Badminton Games with Simple Random Walk

Adity Banerjee, Ankit Bhar, Anunoy Chal,
Aryan Sahu, Aryanil Kundu, Atmadeep Sengupta

1 Content

2 Introduction

In Statistics and probability, there is a interesting topic called simple symmetric random walk.The concept is each time you toss a unbiased coin to decide the direction of your next move.($i.e$ the probability of going right and left are both 0.5).

Here we want to verify actually if those mathematical stuff is followed by the badminton games' results or not in various way. Based on that analysis, we want to create a betting strategy for that particular game.

Assumption And About the Data: A mentionable assumption , to handle the symmetric random walk we are presuming that the two player are of equal ability and every point played are independent of each other.And thus we have collected the scores of those games which have been finished $19-21$, $19-21$ i.e the close games. For the badminton games which goes to deuce (beyond the scores 21), we have truncated our data upto $20-20$.

Further that, we will try to establish whether the calculated probability of a walk follows the proportion of a particular score (e.g. proportion of scores $21-5$) in the data set or not.

Source: The data (*Badminton BWF World Tour games from 2018 to 2021*) has been collected from *Kaggle*.

3 Fitting positive sojourn time with close games

Here, we want to fit badminton games with score $20-20$ with simple random walk. We will check the time in lead for a player and fit it with the theoretical probabilities of positive times. At first we want to get the theoretical probabilities.

Equidistribution Theorem:
The probability that paths of length 2n such that $S_{2n}=0$ has exactly 2k of its sides above the axis is independent of k and equal to $\frac{u_{2n}}{n+1}$ (Here k=$0,1,\ldots,n$)

Some notations which will be used in the proof :

$$u_{2n} = P(S_{2n}=0) = \frac{\binom{2n}{n}}{2^{2n}}$$

$$f_{2n} = P(S_1 \neq 0, S_2 \neq 0, \ldots, S_{2n-1} \neq 0, S_{2n}=0)$$

Proof :

Consider the case k=0 and k=n . The number of paths to (2n,0) with all sides above the x axis is equivalent to number of paths from (1,1) to (2n,0) without touching the line directly below the x-axis . Now using reflection principle the number of paths is the difference between the number of paths from (1,1) to (2n,0) and the number of paths from (1,-3) to (2n,0) i.e.

$$\binom{2n-1}{n} - \binom{2n-1}{n+1} = \frac{\binom{2n}{n}}{n+1}$$

This proves the assertion for k=n . By symmetry it is also true for k=0 . Now by induction we assume it true for $1 \leq k \leq n$. Now we will apply induction over n . The theorem is easily verified when n = 1 . Assume it is true for all paths of length less than 2n . Now we have to show it for path of length 2n . Let 2r be the epoch of first return to x axis. Now there are two cases :

Case 1 : The section of the path upto epoch 2r is above the x axis . Then we must have $1 \leq r \leq k$ and the remaining path of length 2n-2r must contain a length of 2k-2r on the positive side . For the first section of path of length 2r the number of ways are $2^{2r-1} f_{2r}$ And for the rest path of length 2n-2r where exactly 2k-2r amounts of sides are above the x axis ,the number of such paths are $\frac{2^{2n-2r} u_{2n-2r}}{n-r+1}$ by equidistribution theorem which is already assumed true for path of length less than 2n . So the total number of paths in this case are

$$\frac{2^{2r-1} \cdot f_{2r} \cdot 2^{2n-2r} \cdot u_{2n-2r}}{n-r+1}$$

Since r varies from 1 to k then total number of paths of case 1 :

$$\sum_{r=1}^{k} \frac{2^{2r-1} \cdot f_{2r} \cdot \binom{2n-2r}{n-r}}{2(n-r+1)} = \sum_{r=1}^{k} \frac{\binom{2r-2}{r-1} \cdot \binom{2n-2r}{n-r}}{r(n-r+1)}$$

[Here we are using the fact that $f_{2n} = \frac{u_{2n}}{2n-1}$]

Case 2: The path upto epoch 2r lies totally below the x axis and there are exactly 2k sides above the x axis in the remaining 2n-2r length of path . Then we must have $k \leq n-r$ .Now using similar argument as in the case 1 the total number of paths here will be

$$\sum_{r=1}^{n-k} \frac{2^{2r-1} \cdot f_{2r} \cdot \binom{2n-2r}{n-r}}{2(n-r+1)} = \sum_{r=k+1}^{n} \frac{\binom{2r-2}{r-1} \cdot \binom{2n-2r}{n-r}}{r(n-r+1)}$$

[In the sum we are changing the index from r to n-r+1] Now r varies from k+1 to n and the terms under summation are equivalent to case 1 .
From case 1 and case 2 it follows that number of paths with 2k positive sides is

$$\sum_{r=1}^{n} \frac{\binom{2r-2}{r-1} \cdot \binom{2n-2r}{n-r}}{r(n-r+1)}$$

which is independent of k . Since the total number of paths is $2^{2n} u_{2n}$ and the there are n+1 possible values of k So the required number of paths $\frac{2^{2n} \cdot u_{2n}}{n+1}$ Hence, the required probability is $\frac{2^{2n} \cdot u_{2n}}{2^{2n}(n+1)} =$

$$\boxed{\frac{\cdot u_{2n}}{n+1}}$$

Now, the conditional probability that, the run remains $2k$ time positive, given that $S_{2n} = 0$ is

$P(2k \text{ times positive } | S_{2n} = 0) = \frac{P(2k\text{times positive} \cap S_{2n}=0)}{P(S_2 n=0)} = \frac{\frac{u_{2n}}{n+1}}{u_2 n} =$

$$\boxed{\frac{1}{n+1}}$$

Now, setting n=20, we want to fit with badminton games with 3σ limits.

Hence, games with score are fitting with theoretical probabilities.

Now, we want to see data points of close games, even if not draw, fits with theoretical probability or not. At first, we find the probabilities.

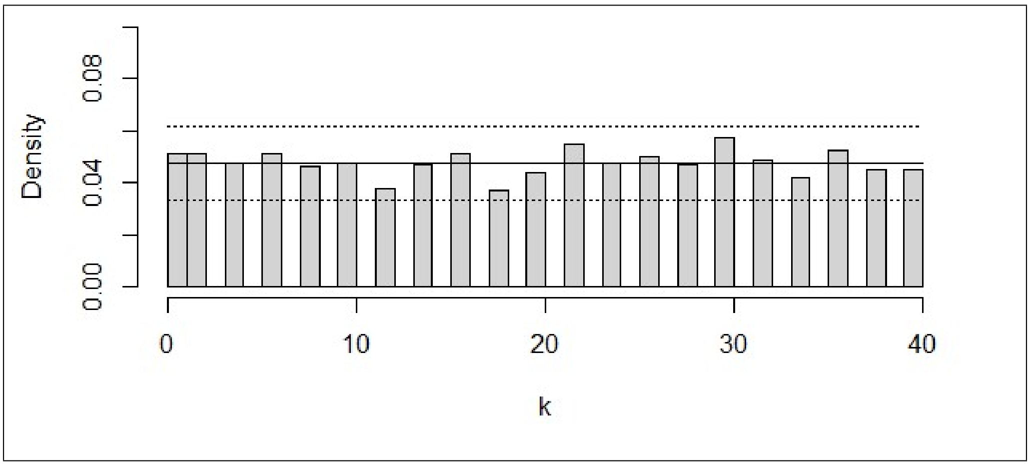

Figure 1: positive times of 20-20 game

Problem:

Consider a badminton game which ends at time 2a, for some positive integer a. Note that this 2a is basically the sum of the final scores of the two players in the badminton game. For this case, assume that this game has not ended in a draw. Suppose the absolute difference between the final scores is $2b(b > 0)$. Fix any one of the players .If that player has won, let 2m, $(b \leq m \leq a)$ to be the amount of time this player was in lead till time 2a and similarly consider 2m,$(0 \leq m \leq b - a)$, to be the amount of time this player was in lead till time 2a if the player has lost the game. Calculate the probability of getting a game of $2a$ points where absolute difference of points of two players is $2b(b > 0)$ points and was leading by $2m$ points.That means probability of getting a path from $(0, 0)$ to $(2a, 2b)$ with $2m$ time positive. Also calculate the conditional probability of these positive sojourn times given that the path ends at time 2a.

Solution:

Let us consider paths starting from the origin until time $2a$, where $a \in \mathbb{N}$. Let the two players be Player A and Player B. Now, let us focus on any one of the players, say Player A. Assuming that this game has not ended in a draw, we need to consider two cases:

1. Player A wins,
2. Player A loses.

For the first case, A wins by a score of $2b$, $0 < b \leq a$. We are going to consider the amount of time Player A was in the lead, which corresponds to ***paths till*** $(2a, 2b)$ ***with*** $2m$ ***amount of time above the x-axis, where*** m ***varies from*** b ***to*** a***.***

So now let us try to find out the number of such paths. For this purpose, we will consider their last visits to the origin. Let the last visit to the origin happens at time $2k$. We need to know the bounds on k. Clearly, $2m - (2a - 2k)$ should be greater than or equal to 0, which gives $k \geq a - m$ as the path remains in positive side from $(2k, 0)$ to $(2a, 2b)$. Also, k can be at most $(a - b)$ as path should be in positive side for last b times. So, combining both bounds, we get $a - m \leq k \leq a - b$.

Now, after time $2k$, the path should never touch or cross the x-axis until $(2a, 2b)$. To calculate this, we will take the total number of paths from $(2k, 0)$ to $(2a, 2b)$ and subtract those paths that touch or cross the x-axis. The number of paths from $(0, 0)$ to (n, r) is given by $N_{n,r}$, which is equal to $\binom{n}{\frac{n-r}{2}}$, where n and r have the same parity. Now, from $S_{2k} = 0$, the path will go to $S_{2k+1} = 1$ by our conditions. So, the total number of paths from $(2k + 1, 1)$ to $(2a, 2b)$ is the same as the number of paths from $(0, 0)$ to $(2a - 2k - 1, 2b - 1)$, which is equal to $\binom{2a-2k-1}{a-b-k}$. Our next task is to calculate the number of paths from $(2k + 1, 1)$ to $(2a, 2b)$ which touch or cross the x-axis. Since we are dealing with symmetric random walks, by using the Reflection Principle, we can take the reflection of the point $(2k + 1, 1)$ about the x-axis and say that the paths from $(2k + 1, 1)$ to $(2a, 2b)$ have a bijection with the paths from $(2k + 1, -1)$ to $(2a, 2b)$. These paths are bound to cross the x-axis to reach $(2a, 2b)$, as $b > 0$. So, our resulting number of paths equals $\left[\binom{2a-2k-1}{a+b-1-k} - \binom{2a-2k-1}{a+b-k}\right]$.

Now we are ready to calculate the number of paths from the origin to $(2a, 2b)$ with $2m$ amount of time above the x-axis by taking their last visit to the origin at some time $2k$, which is as follows:

$$\boxed{\sum_{k=a-m}^{a-b}\left[\frac{\binom{2k}{k}\left(\binom{2a-2k-1}{a-b-k}-\binom{2a-2k-1}{a-b-k-1}\right)}{2^{2a}(k+1)}\right]}$$

When $m = 0$, we will take the sum from 0 to $(a-1)$. Therefore, using the result we derived above, we will calculate the conditional probability:

$$P(2m \text{ times positive}|S_{2a} = 2b) = \frac{P((2m \text{ times positive}) \cap (S_{2a} = 2b)}{P(S_{2a} = 2b)}$$

So, the above probability becomes:

$$\boxed{\frac{\sum_{k=a-m}^{a-b}\left[\frac{\binom{2k}{k}\left(\binom{2a-2k-1}{a+b-1-k}-\binom{2a-2k-1}{a+b-k}\right)}{2^{2a}(k+1)}\right]}{\binom{2a}{a+b}}}$$

In this way, we will get the probabilities for each $m = a-b, a-b+1, \ldots, a$.

Similarly, we can work on the case in which Player A loses the game by a score of $2b$, $0 < b \leq a$. In this case, we will consider the paths from $(0,0)$ to $(2a,-2b)$. Here we will focus on the amount of time in which Player A was behind. So, the problem becomes:***number of paths from*** $(0,0)$ ***to*** $(2a,-2b)$ ***with*** $2a-2m$ ***amount of time below the x-axis, where*** m ***varies from*** 0 ***to*** $(a-1)$. Then it remains above the x-axis for $2m$ amount of time.

So, by reflection, we can get a bijection of the paths from $(0,0)$ to $(2a,-2b)$ and $2a-2m$ times in negative side with the paths from $(0,0)$ to $(2a,2b)$ and $2a-2m$ times in positive side. For this paths, if last visit in origin is at time $2k$, then $a-(a-m) \leq k \leq a-b$, i.e. $m \leq k \leq a-b$

$$\sum_{k=m}^{a-b}\left[\frac{\binom{2k}{k}\left(\binom{2a-2k-1}{a-b--k}-\binom{2a-2k-1}{a-b-k-1}\right)}{2^{2a}(k+1)}\right]$$

Therefore, using the result we derived above, we will calculate the conditional probability:

$$P(2m \text{ times positive}|S_{2a} = -2b) = \frac{P((2m \text{ times positive}) \cap (S_{2a} = -2b)}{P(S_{2a} = -2b)}$$

So, the above probability becomes:

$$\boxed{\frac{\sum_{k=m}^{a-b}\left[\frac{\binom{2k}{k}\left(\binom{2a-2k-1}{a+b-1-k}-\binom{2a-2k-1}{a+b-k}\right)}{2^{2a}(k+1)}\right]}{\binom{2a}{a+b}}}$$

In this way, we will get the probabilities for each $m = 0, 1, \ldots, (a-b)$.

Now, we set $a = 20$, $b = 1$ and $b = -1$, and calculate the conditional probabilities. Then, from our dataset, we sort the corresponding datapoints and fit them with the probabilities with 3σ limits.

So, our data fits with the theoretical probabilities and hence we conclude that badminton games fit with simple random walk, atleast for close games!

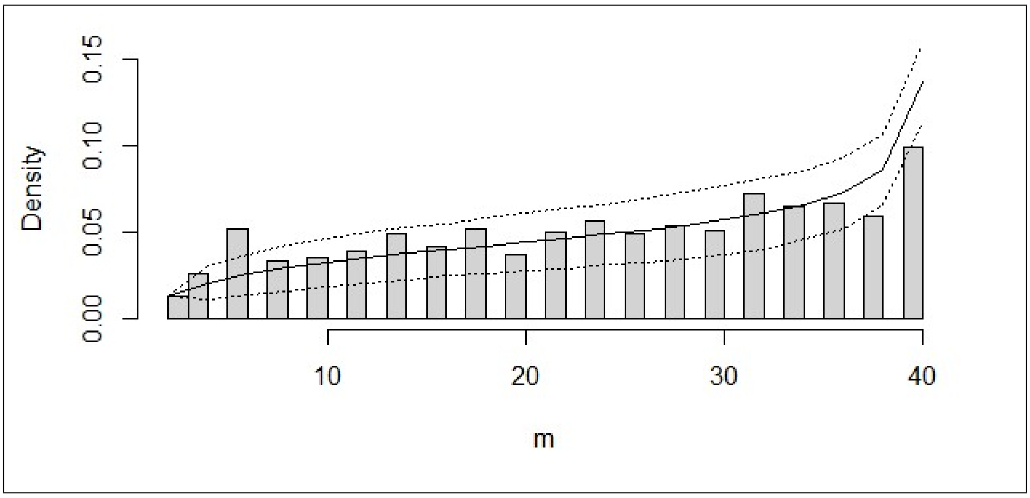

Figure 2: games with score 21-19 and winner leads for 2m times

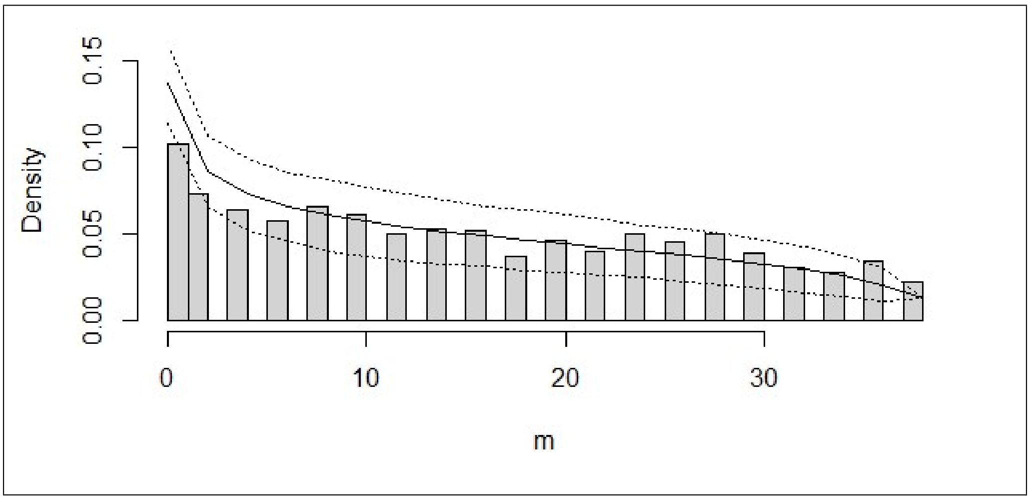

Figure 3: games with score 19-21 and winner leads for 2m times

4 Fitting game results with simple random walk

Now, let's examine data points from a complete game (i.e. a player win with a score of 21) and compare it with theoretical probability calculated by assuming it as a simple random walk. We will start by calculating the theoretical probability.

Problem: Find the probability that a badminton game ends at the score $21 - x$ ($x \in \{0, 1, \ldots 20\}$)

Solution: Let's analyze the situation. In a badminton game, the first player to reach a score of 21 points wins, if we stop the game at that point and not allow deuce. If the game ends with a score of $21 - x$, it means Player A has scored 21 points and Player B has scored x points, (where $x \in \{0, 1, \ldots 20\}$) or vice versa.

Now, let's consider the probability of Player A winning the game with a score of $21-x$. At each point, there are two possible outcomes: either A wins the point or B wins the point. Since both players have an equal chance of winning a point, the probability of A winning a single point is $\frac{1}{2}$. Probability that A wins with score $21 - x$ is $\frac{1}{2} \times \frac{\binom{20+x}{x}}{2^{20+x}}$ as we can choose x points where player B won from the first $20 + x$ points, and player A won the last point. Furthermore, player B wins with the score $21 - x$ with the same probability, we arrive at the required probability =

$$\boxed{\frac{\binom{20+x}{x}}{2^{20+x}}}$$

.
Now, we take our data and fit with the theoretical probabilities.

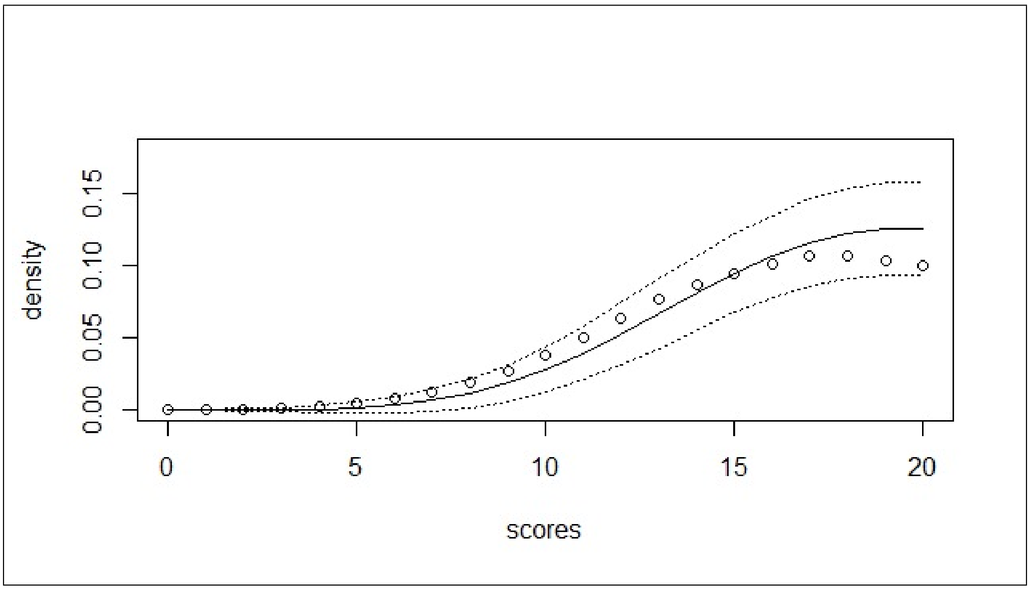

Figure 4: scores of badminton games

So, we conclude from this fitting that simple random walk just not only fits with close games, but in any game in world standard! So, we can model anything on badminton game scores just by simulating random walks. Here we present such a betting strategy.

5 Betting strategy

Problem:

Let there is a betting game on a badminton game where the bettor will predict the score of the game (irrespective of winner). e.g., he or she will bet on the score $21-5$ or $21-19$. Whenever the game goes on after deuce, we will just consider the game until any one of the player gets 21 points. And after the half time,(after any one player reaches 11 points first), the bettor can change his or her bet seeing the game trend, but in the cost of some fine. Let us assume that the bettors bet on the scores in the proportion of theoretical probability of winning, i.e., if probability of getting a score is p, then a person will bet on that score with probability p. So, if a person bets with capital n rupees, he or she gets $\frac{n}{p}$ rupees. If he or she wants to change his or her bet, have to give a fine of en rupees. Find the optimum strategy for giving(or not giving) the fine.

Solution:

Suppose a person bets with capital n rupees on score $21-x$ and at first his expectation of gain is $p_x \frac{n}{p_x} - n = 0$ rupees where p_x is the probability of getting the score $21-x$(assuming the game follows simple random walk). Let at half time the scoreline is $11-y$. Now we will consider two scenarios, we will find the expected gain of the person in both the cases if he changes the bet or not at half time. Define $q_{a,b}$ = probability of the score being $21-b$ from $11-a$ at half time.

<u>*Case I:*</u> Person changes his bet at halftime from $21-x$ to $21-z$. So $q_{y,z}$ = probability of score being $21-z$ from $11-y$. In this case the expectation of gain is $\frac{n}{p_z} q_{y,z} - en - n$ (en is substracted for fine)

<u>*Case 2:*</u> Person did not change his bet and keeps it to $21-x$. So now the expectation of gain is $\frac{n}{p_x} q_{y,x} - n$
Now we want to find that value of z which maximises the gain of the person. Clearly a person changes his bet for more gain, so comparing the gains in the 2 cases, i.e. the person changes the bet when

$$\frac{n}{p_x} q_{y,x} - n < \frac{n}{p_z} q_{y,z} - en - n \implies \boxed{\frac{q_{y,x}}{p_x} < \frac{q_{y,z}}{p_z} - e}$$

This inequality is in terms of y and z and for each value of y we want to find that value of z which maximises the expectation of gain.
Define $f(y) = max\{\frac{q_{y,z}}{p_z} | z = 0, 1, 2, \ldots, 20\}$. Define $g(y) = z$ where$f(y) = \frac{q_{y,z}}{p_z}$ (basically $g(y)$ denote that value of z which maximises $\frac{q_{y,z}}{p_z}$). Now we can change our bet from $21-x$ to $21-g(y)$.

Now, $\boxed{p_x = 2 \times \frac{\binom{20+x}{20}}{2^{21+x}} = \frac{\binom{20+x}{20}}{2^{20+x}}}$ as, we can chose 20 places for first 20 points of the winner among the $20+x$

points played and also choosing the winner in $\binom{2}{1}$ ways. And, to find $q_{y,x}$ we consider three cases.

<u>*Case 1:*</u> $y > x \implies \boxed{q_{y,x} = 0}$ as none of the player can achieve the score x.

<u>*Case 2:*</u> $y < x < 11$. Then only the player trailing can achieve the score x. So total points needed to be played $= 21 + x - 11 - y$=$10 + x - y$. Among which the player in lead will get the last point and 9 more points in between. So, the probability becomes

$$\boxed{q_{y,x} = \frac{\binom{10+x-y-1}{9}}{2^{10+x-y}} = \frac{\binom{9+x-y}{9}}{2^{10+x-y}}}$$

<u>*Case 3:*</u> $y < 11 \leq x$. Then, both the players can get x points.
<u>*Subcase 1:*</u> The player trailing gets x points. Then similar to previous case, the probability is $\frac{\binom{9+x-y}{9}}{2^{10+x-y}}$.
<u>*Subcase 2:*</u> The player leading gets x points. So, winning player gets the last point, and $20-y$ points in between. So, for this subcase, probability becomes $\frac{\binom{9+x-y}{20-y}}{2^{10+x-y}}$
So,

$$\boxed{q_{y,x} = \frac{\binom{9+x-y}{9} + \binom{9+x-y}{20-y}}{2^{10+x-y}}}$$

Now we construct a indicator matrix $((a_{ij}))$ for a fixed e made of $1's$ and $0's$ where $(y+1, x+1)$ th entry denotes that whether we should change our bet at half time or not.

- If $a_{(y+1)(x+1)} = 1$ then we will change our bet from $21 - x$ to $21 - g(y)$
- If $a_{(y+1)(x+1)} = 0$, then we will not change our bet at half time

For example, if we take $e = 0.1$, our indicator matrix will be following.

$$\begin{bmatrix}
0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}$$

Now, we want to see if this betting strategy actually helps or not. As we can see from our study that world class badminton games do follow simple random walk, so we just simulate matches and apply our betting strategy in the simulated games.We have used our betting strategy 1000 times and getting much more gain (63% of the expense in a simulation) in most of the simulations. Whereas, the organizer of the betting also gains in a game by this rule of fine(7% in a simulation) even bettors uses the strategy. So, both ways it generates gain. Hence, it is a nice game to play!

6 Conclusion

- We have showed how the conditional verification should be done.
- Comparing between the collected data and simulating data and measuring the error part we can conclude that the close badminton game follows simple random walk.
- Then, we have checked that not only closed game but any game with score $21 - k$ (where $k \in \{0, ..., 20\}$) follows simple random walk in the sense of proportion of getting the score $21 - k$ matches with the theoretical probabilities.
- After that, we can come with a betting strategy can be created based on the walk,which has been done.

Which has established the relation between Badminton game and simple random walk.

7 Acknowledgement

We acknowledge with a deep sense of reverence ,our gratitude towards our Statistics teacher ***Prof. Arnab Chakraborty*** for his incessant guidance and valuable suggestions given to us in completing this project. Alongwith we will be grateful to acknowledge all the helps from the book *An Introduction to Probability Theory and its Applications, Vol I* by ***William Feller***.