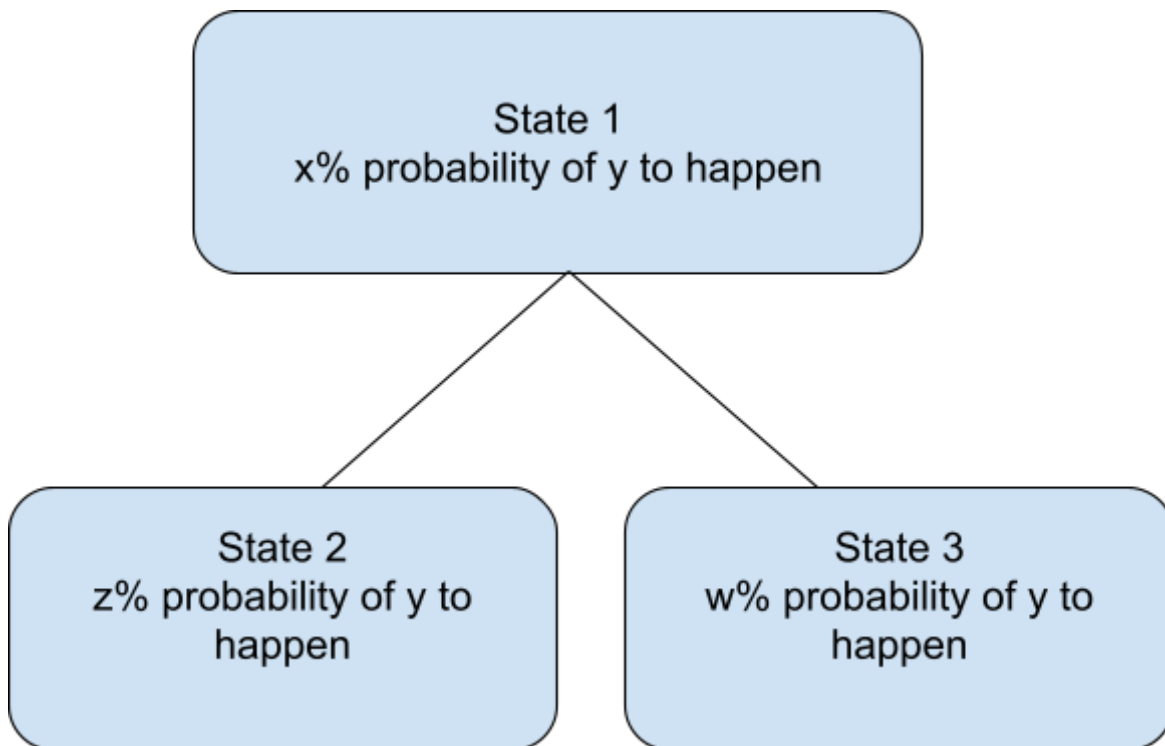


Markov Chain in Sports

1. What is a Markov Chain?

A Simple Markov Chain is a predictive model that determines the probability of a future event based solely on the current state.



Markov chains are used to analyze the probability of the event “y” and the changes to the probability of the event “y” upon the occurrence of Event 1 , 2 ,3 ...

2. Markov Chains in Sports

We can use this in sports like volleyball , tennis , badminton , cricket ,etc. to analyze certain actions, known as states in this chain, like a pass in volleyball , a rally in tennis or badminton , or a certain type of shot in cricket that leads to a favorable outcome , like a point in a net sport , or a boundary in cricket. This helps teams analyze which states are most effective in certain circumstances allowing them to effectively use or counter them. However, doing this in football , where there is no definite number of actions in a play that can bring a guaranteed end

to a play, is much harder (excluding full time and half time). For example a match in football can have almost a thousand actions , like passing, pressing , tackling, shooting , dribbling , making runs , etc. without a guaranteed goal. Meanwhile cricket is guaranteed to have a dot ball , 4 , 6 , running , or a wicket in every delivery .Thus making plays easy to analyze as the outcomes can be easily listed for each action meaning there are a limited number of states.

3. Terms related to the Simple Markov Chain

A **Simple Markov Chain** is a system where the probability of moving to the next state depends only on the current state, not the past history.

A **state** is a situation or case rather,

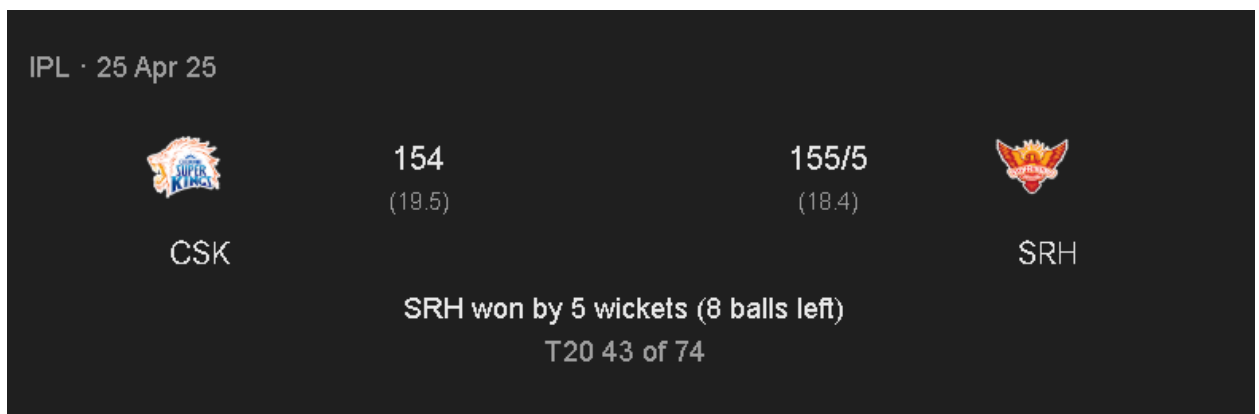
A **transition** is defined as the movement between two states,

Probability in this context refers to the chance of a certain transition to occur,

4. Building a model

To build our very own Cricket Model

We will analyse the deliveries in an **Indian Premier League Cricket Game**, between the **Chennai Super Kings** and **Sunrisers Hyderabad** , dated **25th of April 2025**, Won by **Sunrisers Hyderabad** by **5 wickets** with **8 balls left**



*All data has been taken from *'Match data - Cricsheet'*

4.1 Defining the States in our model

We can make a markov chain by defining the result of each ball as 4 states-

1.**Dot** – A ball in which the batter is unable to score any runs

2. **Wicket**– A ball that results in the batter getting out

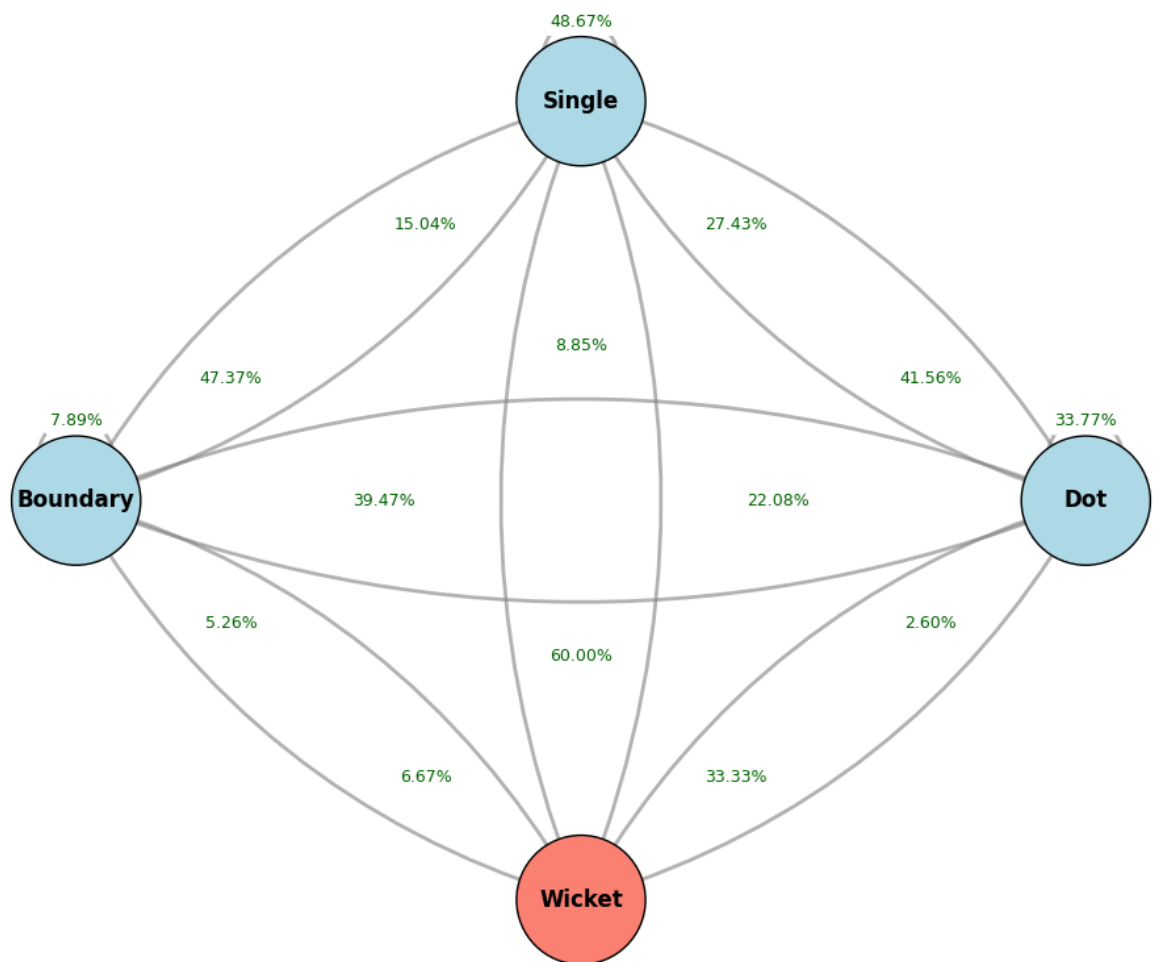
3. **Single** – Used interchangeably with a ball that results in a run , in the case of this model, therefore even if the batters run twice , for simplicity sake we'll consider it a “single”

4. **Boundary** - A ball in which the batter scores either a ‘4’ or a ‘6’

4.2 A graphical representation of a Markov Chain

These states are represented as circles in the figure given below , the lines that are drawn beneath them are simply containers for the values in green which represent the chance (in %) of one state moving to another , the percentage written on top of certain circles represent the chance that the state occurs again on the next ball.

Cricket Markov Chain



Google collab link for the python code of this graph - [Untitled2.ipynb](#)

5. The Matrical representation of our Model

Another representation of the model can be in the form of a matrix/ table as shown below:

From→ To	Dot	Single	Boundary	Wicket
Dot Ball (0 runs)	33.80%	41.60%	22.10%	2.60%
Single (1 - 3 runs)	27.40%	48.70%	15.00%	8.90%
Boundary (4+ runs)	39.40%	47.80%	7.90%	5.30%
Wicket (Out ball)	33.30%	60.00%	6.70%	Negligible

6. What this Data tells us

6.1 Singles are the most dangerous state for the bowling side

The chance for a **wicket (8.9%)** after a single is more than **three** times the **chance** of a wicket to occur after a **dot ball (2.6%)**. This is one of the more *counterintuitive* findings in the data set, despite dot balls creating more **psychological pressure** on the batter and leaking less runs, in most cases batters who come into play right after rotating a strike are off balance or unsettled to say the least. I believe that this causes them to often crumble to the pressure and either get forced into playing defence , something that only increases pressure in T20's or try too hard to force a boundary and end up losing their wicket.

6.2 Singles also cycle into themselves

The other side of the coin however, reveals a very different story once a single is made, the chance for it to “cycle ” into another single is significant (48.7%). In other words Singles tend to feed even more singles. Meaning that they are statistically the most stable and productive action a batting duo can take. Take the example of **AB De Villiers** and **Virat Kohli**, their understanding of the **cyclic nature of singles**, while they played together for the IPL franchise Royal Challengers Bangalore , was so great that they are known as one of the best batting duos in the **history** of the T20 format.

6.3 Boundaries are the hardest to sustain

On the contrary, boundaries are the **least sustainable batting action**, with the chance of consecutive boundaries being lower than the chance of consecutive singles and dot balls. Therefore the aftermath of hitting a boundary is far more likely to be a single or a dot ball rather than another boundary. Henceforth rendering what is meant to be the pinnacle of the T20 game as statistically impractical to rely on. While there are of course statistical anomalies to this rule such as Rohit Sharma , Chris Gale, Harry Brook and Glenn Maxwell who are noted for their boundary frequency, this statistical trend appears to govern the remainder of the T20 game.

6.4 New batters are initially cautious

The data *suggests* that new batters **tend** to be cautious right after a wicket as, the data shows us that boundaries tend to be incredibly rare immediately after a wicket with there being a miniscule 6.70% chance of a boundary to occur after a wicket. This therefore suggests that new batters prefer to play safe right after a wicket, perhaps to avoid the embarrassment of leaving the field early. However this inference often does not orient with reality due to the nature of real life IPL games, in which every ball matters, due to which even lower order batters such as Andre Russell have made a name for themselves as fearsome finishers, aiming to hit boundaries from the get-go. This method has amassed Andre Russel over a 100 sixes in the 120 ball format.

7. Predictive Application of Markov Chains

So far in this essay we have built our own markov chain and analyzed the outcomes and changes between states. However, the true essence of Markov Chains lies in their predictive power, henceforth we will now build our own predictive model using our previous data.

We can predict the result of the next ball with considerable accuracy using our model. The model works on the principle of simple markov chains, simply taking the current state, looking at the data and bringing up the probabilities: As seen here :

[Click the text for Demo](#)

8. Conclusion

In conclusion, The Markov Chain is a mathematical concept with great analytical and predictive potential, especially when utilized in sports such as cricket. I have thus proved that the markov chain is capable of bringing data based, rational strategies to sports such as cricket .