

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

Escuela de Ingeniería y Ciencias

Ingeniería en Ciencia de Datos y Matemáticas

## Cadenas de Markov

Escenario PBL3

ANÁLISIS DE MÉTODOS DE RAZONAMIENTO E INCERTIDUMBRE

*José Miguel Pérez Flores* A00832401

*Kevin Montoya Campaña* A01740352

*José Pablo Sánchez González* A01412539

*Ricardo de Jesús Balam Ek* A00831262

*Carlos Mateos Pérez* A01654085

**Supervisado por**  
Marco Otilio Peña Díaz

Monterrey, Nuevo León. Fecha, October 10, 2022

# 1 Problematización

El deporte permanece como una afición para muchos, para otros, una combinación entre el gusto y la oportunidad de alcanzar la libertad económica. Las casas de apuesta y casinos han abierto oportunidades para atender a este tipo de espectadores; tanta ha sido la demanda que para muchas casas estas oportunidades se han convertido en su principal fuente de ingresos, llevándolas a innovar en la manera de apostar en el deporte. Ahora los jugadores pueden apostar cualquier ocurrencia en el deporte, como la cantidad de tiros a puerta que realiza un jugador, o la ruta de partidos que tendrá un equipo de baloncesto.

Los apostadores pueden jugar al juego de la casa y apostar con el corazón, pero difícilmente encontraran el éxito a un largo plazo. No obstante, existen otro tipo de jugadores que juegan con la información para tomar la mejor decisión posible, permisible gracias a la toma de datos y a diferentes tipos de análisis probabilísticos.

Vencer a la casa no es meramente imposible, el jugador se puede respaldar de técnicas y modelos estocásticos para lograr este cometido; para llegar a esto, primero hay que abstraer las matemáticas y teoría que solucionan la problemática.

## 2 Enfoque

Los deportes son una de las formas de entretenimiento mas populares, lo cual también lo ha convertido en uno de los ámbitos mas involucrados con las apuestas. Uno de los intereses mas comunes de los aficionados y los apostadores es intentar predecir en que posición van a quedar los equipos basados en la poca información que tiene. Debido a la gran cantidad de factores que influyen esto, como las lesiones o los vaivenes en el nivel de juego, esto no es una ciencia exacta, por lo que se puede tomar un enfoque probabilístico.

## 3 Propósito

Definir un modelo de cadena de Markov con al menos 6 estados que represente un problema y sus transiciones donde se defina la matriz de transición y distribución y realizar al menos dos pruebas las cuáles consisten en que dado un estado inicial, sacar su vector de probabilidad en un momento  $X$  futuro y definir una secuencia y sacar su probabilidad de acuerdo con el modelo definido. Nuestro modelo busca predecir el éxito de un equipo de la NBA basado en sus resultados previos.

## 4 Información

Las cadenas de Markov, nombradas así por el matemático ruso Andrey Markov, son sistemas matemáticos que saltan de un estado a otro, con dos estados en el espacio (A y B) existen 4 posibles movimientos, ya que el estado A puede moverse al B o regresar al mismo A, en este diagrama de 2 estados la probabilidad de movimiento de un estado a otro es de 0.5 (Ghahramani, 2001).

Para realizar modelos no siempre se dibujan los diagramas de las cadenas, sen vez de eso, se utilizan matrices de transición para contabilizar las probabilidades de transición. Cada estado del espacio es incluido una vez como columna y una como fila, cada celda indica la probabilidad de moverse del estado del columna a su estado de fila. Por lo que en la matriz, las celdas realizan la misma función que una flecha podría realizar en un diagrama.

Las cadenas de Markov pueden llegar a ser bastantes grandes y poderosas en temas meteorológicos, ecologistas, de las ciencias de la computación, ingeniería financiera y algunos otros que necesiten modelar grandes fenómenos (Powell, 2014).

## 5 Razonamiento

Para la generación de la Matriz de transición, se utilizaron los datos históricos de las ultimas 17 temporadas de la NBA. Se decidió este periodo de tiempo debido a que antes de este periodo, la liga tenia menos equipos de los que tiene actualmente. Esto es un método frecuentista de asignar las probabilidades, donde asumimos que la distribución de los hechos en el pasado es igual a la probabilidad real de que estos pasen.

Clasificamos a todos los equipos y les asignamos uno de 7 estados, dependiendo de su resultado en la ultima temporada. Los estados son:

- **Final:** Llego a la Final
- **Final Conferencia:** Eliminado en la Final de Conferencia
- **2da Ronda:** Eliminado en segunda ronda de playoffs
- **1ra Ronda:** Eliminado en primera ronda de playoffs
- **Pelea Playoffs:** Los 4 equipos que quedaron justo afuera de los Playoffs
- **Resto:** Equipos mediocres que no pelearon por llegar a playoffs o por tener el peor record
- **Bottom 4:** 4 equipos con el peor record

	Final	Final Conferencia	2da Ronda	1ra Ronda	Pelea Playoffs	Resto	Bottom 4
Final	0.3824	0.0882	0.2647	0.1765	0.0294	0	0.0588
Final Conf.	0.050	0.1	0.3	0.4	0.15	0	0
2da Ronda	0.1029	0.1324	0.1618	0.3676	0.1324	0.0735	0.0294
1ra Ronda	0.0368	0.0956	0.1765	0.3382	0.0808	0.1765	0.0956
Pelea Playoffs	0.0735	0.0441	0.1324	0.2353	0.2794	0.1029	0.1324
Resto	0	0	0.049	0.2451	0.1863	0.3627	0.1569
Bottom 4	0.0147	0.0147	0.0147	0.0735	0.0735	0.4265	0.3824

Table 1: Matriz de Transición

También realizamos un diagrama de grafos para mostrar gráficamente las posibilidades de pasar de un estado a otro. Debido a la gran cantidad de estados y la alta interconexion entre ellos, lamentablemente no es una visualización tan clara.

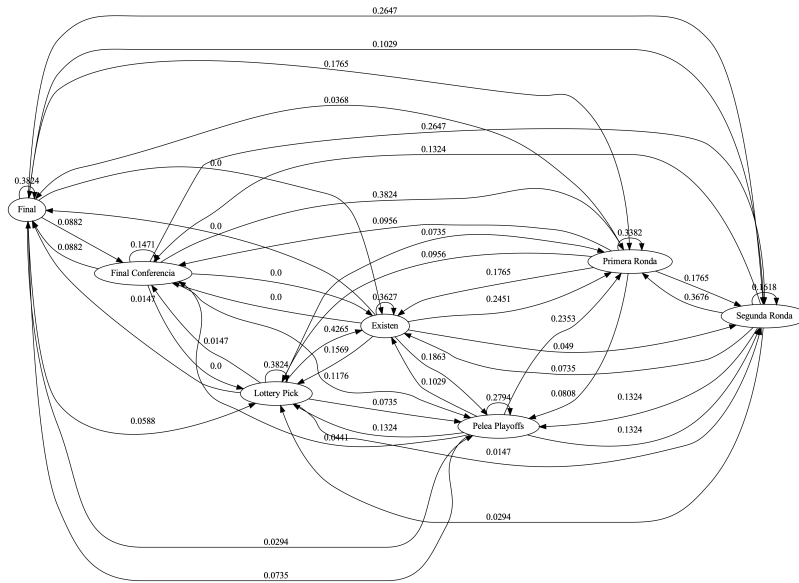


Figure 1: Diagrama de Grafos

Final	Final Conferencia	2da Ronda	1ra Ronda	Pelea Playoffs	Resto	Bottom 4
0.0666	0.0666	0.1333	0.2666	0.1333	0.2666	0.1333

Table 2: Probabilidades estacionarias

## 6 Conclusiones

Si proyectamos los resultados a un tiempo muy grande, las probabilidades de la cadena de Markov se estabilizan en el siguiente vector estacionario.

Este vector no nos da mucha información, debido a que si uno es aficionado de la NBA, rápidamente se da cuenta que simplemente nos indica las proporciones de los equipos que llegan a cada estado. Por ejemplo, 2 equipos de 30 llegan a la final, por lo que esto sería una proporción de 1 sobre 15, que nos da 0.0666. Además, llegamos a este vector estacionario de manera sumamente rápida, tomándonos apenas 20 años para que nuestra cadena nos deje de dar información útil.

Donde si es útil es a corto plazo, donde podemos determinar de mejor manera cuales son las expectativas de un equipo basado en resultados previos. Por ejemplo, si queremos predecir los resultados en 3 años para un equipo que llego a la final. Este caso nos daría el siguiente vector de probabilidad:

Final	Final Conferencia	2da Ronda	1ra Ronda	Pelea Playoffs	Resto	Bottom 4
0.1199	0.0891	0.1736	0.2824	0.1095	0.1309	0.0944

Table 3: Proyección Final en 3 años

Si comparamos este vector con el de las probabilidades estacionarias, nos damos cuenta como un equipo que llego a la final hace 3 años, tiene el doble de probabilidad de llegar a la final que un equipo promedio. También lo podemos comparar con un equipo que hace 3 años estaba entre los 4 peores y veremos que este tiene una distribución de probabilidad completamente diferente.

Final	Final Conferencia	2da Ronda	1ra Ronda	Pelea Playoffs	Resto	Bottom 4
0.0369	0.0405	0.0927	0.2302	0.1450	0.2735	0.1810

Table 4: Proyección Bottom 4 en 3 años

También se pueden usar las cadenas de Markov para determinar las posibilidades de ciertas secuencias. En algunos casos se puede a partir de la frecuencia de una cadena se puede determinar cual es la cadena de Markov adecuada, pero esto se sale del enfoque aplicado al problema actual. En nuestro caso, utilizamos la cadena de Markov generada para predecir que tan improbable es que un equipo vaya a las Finales 4 años seguidos, como famosamente ocurrió con los Golden State Warriors y los Cleveland Cavaliers. Esto se hace multiplicando las probabilidades de cada paso individual, ya que en una cadena de Markov solo recuerda su ultimo estado. En este caso sería la posibilidad de llegar a la Final si estuviste en la Final el año previo elevado a la cuarta potencia, lo que nos da un resultado de 0.0559.

También podemos usar este método para determinar secuencias mas complicadas, como un equipo que se encontraba entre los cuatro peores, que el año siguiente llega a estar en la pelea para clasificar a los playoffs y en su tercer año se elimina en segunda ronda. Aunque esta secuencia suene como algo muy probable y normal, tiene una probabilidad de apenas 0.0097 de ocurrir. Es decir, es mucho menos probable que la posibilidad de un equipo llegando a la final 4 años seguidos.

## References

- Ghahramani, Z. (2001). An introduction to hidden markov model and bayesian networks. hidden markov models: Applications in computer vision. *World Scientific Publishing Co.*, 9–42.
- Powell, V. (2014). Markov chains.