Grado en Estadística Empresarial

Trabajo de Fin de Grado

Curso 2016/2017

Convocatoria de Junio

Modalidad: Investigación.

Título: Cadenas de Markov versus ordenamiento lineal para la unificación de votaciones.

Autor: Arturo Ferrández Ortega.

Tutor: Mercedes Landete Ruíz.

ÍNDICE

1. RESUMEN

En el mundo hay múltiples situaciones donde se necesita realizar un orden o una clasificación dadas unas votaciones o puntuaciones, y hemos considerado interesante realizar un trabajo donde hemos estudiado métodos que nos permiten tomar decisiones sobre cual pueda ser la clasificación o ranking óptimo.

Entre los múltiples métodos que hay, el trabajo se centra en los dos que hemos considerado más importantes, que son Cadena de Markov (Markov Chain) y Ordenamiento Óptimo (Linear Ordering Problem).

Partiremos en ambos métodos de una matriz de datos de votaciones o puntuaciones, la cual debe ser cuadrada, asimétrica y las filas y columnas deben estar en el mimos orden.

Para explicar estos dos métodos y debido a mi motivación personal por enfocarlo a lo que más me gusta que es el deporte, el trabajo se ha centrado en un ejemplo de una clasificación deportiva de los cuatro primeros equipos de la liga de fútbol española.

En dicho ejemplo, primero se planteara el problema, modificando los datos para poder usar cada uno de estos métodos, cuando sea necesario. Se explicará teóricamente cada uno de estos métodos y luego se resolverá de varias maneras, dentro de las cuales se encuentra resolverlo con el software RStudio o Matlab, y por último se compararan los resultados de ambos métodos.

Una parte importante del trabajo, es ver como se aplican estos métodos a casos reales, por ejemplo al Ordenamiento de los Caballos del Vino de las fiestas de Caravaca de la Cruz, donde se aplicaran ambos métodos y se compararan los resultados.

Por último, se explicaran las ventajas y desventajas de cada método, para según las circunstancias, aplicar uno u otro.

2. INTRODUCCIÓN

Como unificación de votaciones entendemos elegir entre distintas alternativas basándonos en lo que dicen los votantes. El acto de votar más popular son las elecciones políticas, en las que se elige de entre los distintos candidatos. Pero las votaciones también pueden utilizarse para elegir entre diferentes planes de acción, conceder premios.

Si solo hay dos opciones, o sólo hay que elegir una de una lista, normalmente se decide por mayoría, es decir si más de la mitad de los votantes se decantan por una de las opciones, ésta debe ser el resultado único de la elección. Podemos decir que es el sistema de votación más sencillo. Pero muchas votaciones requieren reglas más complejas, especialmente si hay más de dos resultados posibles.

Situaciones cotidianas en las que "muchos" opinan son:

- Fútbol: En todas las ligas se recoge información sobre los equipos, y basándose en ella se puede elaborar posibles rankings futuros sobre como ordenar a los equipos.
- Internet: Cada vez que realizamos una búsqueda en Google, y el buscador nos devuelve las respuestas en un determinado orden.
- Programas de televisión: El "prime time", o lo que es lo mismo, que hora es la idónea para emitir los programas, series, películas…, y que estos tengan máxima audiencia.
- Elecciones: Como posicionar a los candidatos de cara a unas elecciones, mediante por ejemplo encuestas de opinión.
- Carreras: Por ejemplo de caballos, donde tenga que votarse por el caballo ganador.
- Precios productos: Si queremos decidir el precio idóneo de un producto, los consumidores pueden opinar sobre lo que estrían dispuestos a pagar por él.

Es frecuente tener que tomar una única decisión a partir de una serie de opiniones o resultados, es como si cada cliente diera su opinión acerca de una ordenación, y como no se van a poner de acuerdo, usamos algún método para obtener un ordenamiento óptimo.

Con este trabajo queremos estudiar los sistemas de votaciones, y la elaboración de distintos tipos de rankings. Para ello vamos a comparar dos técnicas (entre las diversas que hay), que son Cadena de Markov (Markov Chain) y Ordenamiento Lineal (Linear Ordering Problem).

Otro método que se usa para sistemas de votaciones, aunque no lo estudiaremos, es el Método de Condorcet:

- *ORDESHOOK, PETER C.: GAME THEORY AND POLITICAL THEORY. AND INTRODUCTION. Cambridge University Press, 1999.*

  http://asojodcr.blogspot.com.es/2009/03/la-paradoja-de-condorcet.html

Cuando Antoine Caritat Condorcet estaba intentando encontrar cual era el tamaño óptimo de los jurados de la Revolución Francesa justo antes de esta, descubrió un problema matemático en el proceso de votación. Su paradoja nos dice que la transitividad de las preferencias individuales no tiene por qué dar lugar a la transitividad en las preferencias colectivas.

Un método de Condorcet es un sistema de voto que siempre elige el ganador de Condorcet, es decir, el candidato al que los votantes prefieren ante el resto de candidatos, cuando se los compara de uno en uno. El ganador se puede encontrar haciendo series de comparaciones entre dos candidatos. Un sistema de voto que siempre elige el ganador de Condorcet cuando lo hay es un sistema que satisface el criterio de Condorcet.

Vamos a ver ahora un ejemplo ficticio donde podamos aplicarlo. Imaginemos que los habitantes de Andalucía, una Comunidad Autónoma de España, quieren elegir su capital. Las cuatro provincias más pobladas son las candidatas. Imaginemos que todo el electorado quiere que sea la ciudad en la que viven, o que esté lo más cerca posible a su ciudad. Las ciudades candidatas son:

- Sevilla: La ciudad más grande, con el 42% de los votantes pero muy lejana del resto de ciudades.
- Málaga: Con el 26% de los votantes.
- Córdoba: Con el 17% de los votantes.
- Granada: Con el 15% restante.

Las preferencias de los votantes son las siguientes:

42% votantes (Sevilla)	**26% votantes (Málaga)**	**15% votantes (Córdoba)**	**17% votantes (Granada)**
1.Sevilla	1.Málaga	1.Córdoba	1.Granada
2.Málaga	2.Córdoba	2.Granada	2.Córdoba
3.Córdoba	3.Granada	3.Málaga	3.Málaga
4.Granada	4.Sevilla	4.Sevilla	4.Sevilla

Para encontrar el ganador de Condorcet, cada provincia candidata ha de ser enfrentada con el resto de candidatas en batallas imaginarias una a una. En cada emparejamiento, la provincia ganadora es la candidata preferida por la mayoría de los votantes. Los resultados de cada emparejamiento son los siguientes:

Pareja	**Ganador**
Sevilla (42%) vs. Málaga (58%)	Málaga
Sevilla (42%) vs. Córdoba (58%)	Córdoba
Sevilla (42%) vs. Granada (58%)	Granada
Málaga (68%) vs. Córdoba (32%)	Málaga
Málaga (68%) vs. Granada (17%)	Málaga
Córdoba (83%) vs. Granada (17%)	Córdoba

La matriz de resultados es la siguiente:

		A			
		Sevilla	Málaga	Córdoba	Granada
uB	Sevilla		[A] 58% [B] 42%	[A] 58% [B] 42%	[A] 58% [B] 42%
	Málaga	[A] 42% [B] 58%		[A] 32% [B] 68%	[A] 32% [B] 68%
	Córdoba	[A] 42% [B] 58%	[A] 68% [B] 32%		[A] 17% [B] 83%
	Granada	[A] 42% [B] 58%	[A] 68% [B] 32%	[A] 86% [B] 17%	
Resultado final:		4ª	1ª	2ª	3ª

- [A] indica votantes que prefieren a la candidata de la columna a la de la fila.
- [B] indica votantes que prefieren a la candidata de la fila a la de la columna.

Como vemos en los datos anteriores, Málaga gana a cualquier otra candidata cuando se les enfrente una a una. Por tanto, Málaga es la ganadora de Condorcet. Cualquier método de Condorcet daría a Málaga como ganadora. Si hubiéramos usado el sistema de voto por mayoría simple hubiera ganado Sevilla, aunque el 58% de los votantes cree que es la peor opción.

Puede darse el caso de que no exista el ganador de Condorcet. Esto puede ocurrir si hay una especie de empate, una dependencia cíclica. En este caso, la manera de encontrar al ganador de la elección varía entre los métodos de Condorcet. Algunos métodos siguen el procedimiento básico descrito más abajo y definen un método alternativo en caso de no existir el ganador de Condorcet. Otros métodos cuentan los votos de una manera completamente diferente, pero se les considera métodos de Condorcet porque eligen el ganador de Condorcet siempre que este exista.

No en todos los sistemas de votos en los que se escriben en orden de preferencia los candidatos y se elige a uno es posible aplicar el método de Condorcet.

Podemos ver esto de una forma más clara con un sencillo ejemplo ficticio:

Suponemos que existen 7 clientes o individuos {1,2,3,4,5,6,7} que se disponen a elegir entre 3 opciones (solo pueden elegir una), por ejemplo las llamamos {a,b,c}. Eligen por mayoría simple, lo que significa que la opción que se elija debe salir al menos con 3 votos. Además vamos a suponer que se establece una relación transitiva de preferencias para cada individuo como vemos en la siguiente tabla:

	INDIVIDUOS						
ORDEN	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	a	b	c	b	c	b	c
DE	b	a	a	c	b	c	a
	c	c	b	a	a	a	b
PREFERENCIAS	a>b>c	b>a>c	c>a>b	b>c>a	b>c>a	b>c>a	c>a>b

La votación por mayoría simple en este caso demuestra que ninguna opción es socialmente preferible para el grupo de individuos, lo vemos:

La opción a>b por 4 votos frente a 3.

La opción b>c por 4 votos frente a 3.

La opción c>a por 4 votos frente a 3.

Vemos que no hay una sola opción que pueda ganar si las consideramos todas, ya que a vence a b, b vence a c y c vence a a.

Se podría pensar que es casualidad, y que en el mundo real no pasa esto, pero, hay resultados de simulaciones de cuerpos de votantes realizadas por ordenadores referentes a la frecuencia de que ocurran estos ciclos, que demuestran que esto si es probable de que ocurra (porque la probabilidad de ocurrencia de la paradoja es mayor que cero). Además podemos afirmar que la probabilidad de ocurrencia de la paradoja tiende a 1 a medida que vamos aumentando el número de alternativas o de individuos.

Volviendo a los dos métodos objeto de estudio, Cadena de Markov (Markov Chain) y Ordenamiento Óptimo (Linear Ordering Problem), ambos métodos elegidos siempre devuelven una solución:

Como datos de partida tendremos una matriz, esta matriz será cuadrada, pues todos opinan sobre todos, es decir si hay 20 clientes, un cliente opina de los 19 restantes y así todos, luego la matriz será 20 por 20.

En esta matriz las filas y las columnas deben estar el mismo orden, con el fin de que sea más fácil interpretar los resultados.

La matriz además es asimétrica, esto se debe a su naturaleza. Si el individuo i coloca al individuo j en la posición t, el j no tiene por qué opinar que i está en la t.

Decir que si no todos opinan de todos (no es cuadrada la matriz) o fuese simétrica, esta técnicas no sirven.

Para entender mejor en qué consisten estas dos técnicas, vamos a ver un ejemplo donde se detalla paso a paso cada de ellas.

Partimos de un ejemplo ficticio basado en los 4 primeros equipos clasificados de la Liga española. El Barcelona, el Real Madrid, el Atlético de Madrid y el Sevilla. Suponemos los siguientes resultados entre ellos:

EQUIPO 1	RESULTADO	EQUIPO 2
Barcelona	3 - 2	R. Madrid
R. Madrid	2 - 2	Barcelona
Barcelona	1 - 1	Atlético de Madrid
Atlético de Madrid	1 - 0	Barcelona
Barcelona	3 - 1	Sevilla
Sevilla	1 - 2	Barcelona
R. Madrid	2 - 1	Atlético de Madrid
Atlético de Madrid	1 - 0	R. Madrid

R. Madrid	2 -1	Sevilla
Sevilla	3 - 3	R. Madrid
Atlético de Madrid	3 - 2	Sevilla
Sevilla	2 - 1	Atlético de Madrid

Ahora vamos a ver cómo quedaría la clasificación de estos 4 primeros equipos entre ellos según el sistema de puntos establecido en la Liga española, donde cada victoria vale 3 puntos, y en caso de empate se reparte un punto para cada equipo:

Clasificación Normal

PUESTO	EQUIPO	PUNTOS
1º	Barcelona	11
2º	Atlético de Madrid	10
3º	R. Madrid	8
4º	Sevilla	4

A partir de ahora la nueva nomenclatura de los equipos será:

B: Barcelona

M: R. Madrid

AT: Atlético de Madrid

S : Sevilla

Una vez visto cómo quedaría la clasificación por el sistema de puntos de la liga española, vamos a plantear y resolver por los dos métodos de ordenación mencionados anteriormente, y en los cuales se basa este trabajo, Ordenamiento Óptimo y Cadena de Markov, y así ver las posibles diferencias (en caso de que las hubiera) respecto a la clasificación normal.

3. ORDENAMIENTO LINEAL (THE LINEAR ORDERING PROBLEM)

Es un sistema de ordenación que se basa en sumar los elementos por encima de la diagonal principal de una matriz. Se llama Ordenamiento Óptimo, al que da la suma mayor. Este método de ordenar tiene sentido porque se consigue el mayor número de acuerdos posibles.

Ahora construimos las matrices para ver la ordenación óptima del Linear Ordering Problem, basándose en los goles que marca cada equipo (ida y vuelta).

Para saber el ordenamiento óptimo, tenemos que considerar todas las permutaciones posibles, en este caso, como hay 4 equipos, tendremos 4!=24 matrices (24 ordenamientos o permutaciones posibles).

Explicamos un poco la primera matriz, el resto se interpretan igual.

Cada equipo es como si fuese un cliente, y tiene su opinión sobre cuál es el ordenamiento óptimo. En la primera fila, el Barcelona "dice" que el mejor es el Atlético (porque es al que más le cuesta meter goles y por consiguiente ganar, ya que solo le ha metido 1), y Sevilla y Madrid estarían empatados. En la segunda fila el Madrid estaría de acuerdo con el Barcelona, pero en la tercera fila el Atlético ya no está de acuerdo con eso, ya que para él, Madrid y Barça son iguales, y el peor es el Sevilla.

$$\begin{array}{cc} & \begin{array}{cccc} B & M & AT & S \end{array} \\ \begin{array}{c} B \\ M \\ AT \\ S \end{array} & \begin{pmatrix} - & 5 & 1 & 5 \\ 4 & - & 2 & 5 \\ 2 & 2 & - & 4 \\ 2 & 4 & 4 & - \end{pmatrix} \end{array} \quad 22$$

$$\begin{array}{cc} & \begin{array}{cccc} B & M & S & AT \end{array} \\ \begin{array}{c} B \\ M \\ S \\ AT \end{array} & \begin{pmatrix} - & 5 & 5 & 1 \\ 4 & - & 5 & 2 \\ 2 & 4 & - & 4 \\ 2 & 2 & 4 & - \end{pmatrix} \end{array} \quad 22$$

$$\begin{array}{cc} & \begin{array}{cccc} B & AT & M & S \end{array} \\ \begin{array}{c} B \\ AT \\ M \\ S \end{array} & \begin{pmatrix} - & 1 & 5 & 5 \\ 2 & - & 2 & 4 \\ 4 & 2 & - & 5 \\ 2 & 4 & 4 & - \end{pmatrix} \end{array} \quad 22$$

$$\begin{array}{cc} & \begin{array}{cccc} B & AT & S & M \end{array} \\ \begin{array}{c} B \\ AT \\ S \\ M \end{array} & \begin{pmatrix} - & 1 & 5 & 5 \\ 2 & - & 4 & 2 \\ 2 & 4 & - & 4 \\ 4 & 2 & 5 & - \end{pmatrix} \end{array} \quad 21$$

$$\begin{array}{cc} & \begin{array}{cccc} B & S & M & AT \end{array} \\ \begin{array}{c} B \\ S \\ M \\ AT \end{array} & \begin{pmatrix} - & 5 & 5 & 1 \\ 2 & - & 4 & 2 \\ 4 & 5 & - & 2 \\ 2 & 4 & 2 & - \end{pmatrix} \end{array} \quad 19$$

$$\begin{array}{cc} & \begin{array}{cccc} B & S & AT & M \end{array} \\ \begin{array}{c} B \\ S \\ AT \\ M \end{array} & \begin{pmatrix} - & 5 & 1 & 5 \\ 4 & - & 2 & 5 \\ 2 & 2 & - & 4 \\ 2 & 4 & 4 & - \end{pmatrix} \end{array} \quad 19$$

$$\begin{array}{cc} & \begin{matrix} M & B & AT & S \end{matrix} \\ \begin{matrix} M \\ B \\ AT \\ S \end{matrix} & \begin{pmatrix} - & 4 & 2 & 5 \\ 5 & - & 1 & 5 \\ 2 & 2 & - & 4 \\ 4 & 2 & 4 & - \end{pmatrix} \end{array} \quad 21$$

$$\begin{array}{cc} & \begin{matrix} M & B & S & AT \end{matrix} \\ \begin{matrix} M \\ B \\ S \\ AT \end{matrix} & \begin{pmatrix} - & 4 & 5 & 2 \\ 5 & - & 5 & 1 \\ 4 & 2 & - & 4 \\ 2 & 2 & 4 & - \end{pmatrix} \end{array} \quad 21$$

$$\begin{array}{cc} & \begin{matrix} M & AT & B & S \end{matrix} \\ \begin{matrix} M \\ AT \\ B \\ S \end{matrix} & \begin{pmatrix} - & 2 & 4 & 4 \\ 2 & - & 2 & 5 \\ 5 & 1 & - & 5 \\ 4 & 4 & 2 & - \end{pmatrix} \end{array} \quad 22$$

$$\begin{array}{cc} & \begin{matrix} M & AT & S & B \end{matrix} \\ \begin{matrix} M \\ AT \\ S \\ B \end{matrix} & \begin{pmatrix} - & 2 & 4 & 4 \\ 2 & - & 5 & 2 \\ 4 & 4 & - & 2 \\ 5 & 1 & 5 & - \end{pmatrix} \end{array} \quad 19$$

$$\begin{array}{cc} & \begin{matrix} M & S & B & AT \end{matrix} \\ \begin{matrix} M \\ S \\ B \\ AT \end{matrix} & \begin{pmatrix} - & 4 & 4 & 2 \\ 4 & - & 2 & 4 \\ 5 & 5 & - & 1 \\ 2 & 5 & 2 & - \end{pmatrix} \end{array} \quad 17$$

$$\begin{array}{cc} & \begin{matrix} M & S & AT & B \end{matrix} \\ \begin{matrix} M \\ S \\ AT \\ B \end{matrix} & \begin{pmatrix} - & 4 & 2 & 4 \\ 4 & - & 4 & 2 \\ 2 & 5 & - & 2 \\ 5 & 5 & 1 & - \end{pmatrix} \end{array} \quad 18$$

$$\begin{array}{cc} & \begin{matrix} AT & B & M & S \end{matrix} \\ \begin{matrix} AT \\ B \\ M \\ S \end{matrix} & \begin{pmatrix} - & 2 & 2 & 4 \\ 1 & - & 5 & 5 \\ 2 & 4 & - & 5 \\ 4 & 2 & 4 & - \end{pmatrix} \end{array} \quad 23$$

$$\begin{array}{cc} & \begin{matrix} AT & B & S & M \end{matrix} \\ \begin{matrix} AT \\ B \\ S \\ M \end{matrix} & \begin{pmatrix} - & 2 & 4 & 2 \\ 1 & - & 5 & 5 \\ 4 & 2 & - & 4 \\ 2 & 4 & 5 & - \end{pmatrix} \end{array} \quad 22$$

$$\begin{array}{cc} & \begin{matrix} AT & M & B & S \end{matrix} \\ \begin{matrix} AT \\ M \\ B \\ S \end{matrix} & \begin{pmatrix} - & 2 & 2 & 4 \\ 2 & - & 4 & 5 \\ 1 & 5 & - & 5 \\ 4 & 4 & 2 & - \end{pmatrix} \end{array} \quad 22$$

$$\begin{array}{cc} & \begin{matrix} AT & M & S & B \end{matrix} \\ \begin{matrix} AT \\ M \\ S \\ B \end{matrix} & \begin{pmatrix} - & 2 & 4 & 2 \\ 2 & - & 5 & 4 \\ 4 & 4 & - & 2 \\ 1 & 5 & 5 & - \end{pmatrix} \end{array} \quad 19$$

$$\begin{array}{cc} & \begin{matrix} AT & S & B & M \end{matrix} \\ \begin{matrix} AT \\ S \\ B \\ M \end{matrix} & \begin{pmatrix} - & 4 & 2 & 2 \\ 4 & - & 2 & 4 \\ 1 & 5 & - & 5 \\ 2 & 5 & 4 & - \end{pmatrix} \end{array} \quad 19$$

$$\begin{array}{cc} & \begin{matrix} AT & S & M & B \end{matrix} \\ \begin{matrix} AT \\ S \\ M \\ B \end{matrix} & \begin{pmatrix} - & 4 & 2 & 2 \\ 4 & - & 4 & 2 \\ 2 & 5 & - & 4 \\ 1 & 5 & 5 & - \end{pmatrix} \end{array} \quad 18$$

$$\begin{array}{cc} & \begin{matrix} S & B & M & AT \end{matrix} \\ \begin{matrix} S \\ B \\ M \\ AT \end{matrix} & \begin{pmatrix} - & 2 & 4 & 4 \\ 5 & - & 5 & 1 \\ 5 & 4 & - & 2 \\ 4 & 2 & 2 & - \end{pmatrix} \end{array} \quad 18$$

$$\begin{array}{cc} & \begin{matrix} S & B & AT & M \end{matrix} \\ \begin{matrix} S \\ B \\ AT \\ M \end{matrix} & \begin{pmatrix} - & 2 & 4 & 4 \\ 5 & - & 1 & 5 \\ 4 & 2 & - & 2 \\ 5 & 4 & 2 & - \end{pmatrix} \end{array} \quad 18$$

$$\begin{array}{cc} & \begin{matrix} S & M & B & AT \end{matrix} \\ \begin{matrix} S \\ M \\ B \\ AT \end{matrix} & \begin{pmatrix} - & 4 & 2 & 4 \\ 5 & - & 4 & 2 \\ 5 & 5 & - & 1 \\ 4 & 2 & 2 & - \end{pmatrix} \end{array} \quad 17$$

$$\begin{array}{cc} & \begin{matrix} S & M & AT & B \end{matrix} \\ \begin{matrix} S \\ M \\ AT \\ B \end{matrix} & \begin{pmatrix} - & 4 & 4 & 2 \\ 5 & - & 2 & 4 \\ 4 & 2 & - & 2 \\ 5 & 5 & 1 & - \end{pmatrix} \end{array} \quad 18$$

$$\begin{array}{cc} & \begin{matrix} S & AT & B & M \end{matrix} \\ \begin{matrix} S \\ AT \\ B \\ M \end{matrix} & \begin{pmatrix} - & 4 & 2 & 4 \\ 4 & - & 2 & 2 \\ 5 & 1 & - & 5 \\ 5 & 2 & 4 & - \end{pmatrix} \end{array} \quad 19$$

$$\begin{array}{cc} & \begin{matrix} S & AT & M & B \end{matrix} \\ \begin{matrix} S \\ AT \\ M \\ B \end{matrix} & \begin{pmatrix} - & 4 & 4 & 2 \\ 4 & - & 2 & 2 \\ 5 & 2 & - & 4 \\ 5 & 1 & 5 & - \end{pmatrix} \end{array} \quad 18$$

Arriba y a la derecha de cada matriz esta la suma de los elementos que están por encima de la diagonal principal.

Como curiosidad observamos que el mínimo es 17 y coincide con la clasificación donde el Sevilla quedaría en primer lugar, seguido de Real Madrid, Barcelona y Atlético de Madrid (en ese orden), y también con la clasificación donde el primero sería el Real Madrid, en segunda posición estaría el Sevilla, después el Barcelona, y en último lugar el Atlético. Estas clasificaciones ería las más remotas, estrían a la cola de la clasificación óptima.

La clasificación normal que hemos visto anteriormente con el sistema de puntuación de la liga española nos daría una suma de 22 puntos, no sería el ordenamiento óptimo, puesto que hay una clasificación cuya suma de los elementos por encima de la diagonal principal da mayor, en concreto el máximo es 23, esta es la clasificación óptima, que es la que se muestra a continuación:

PUESTO	EQUIPO
1º	Atlético de Madrid
2º	Barcelona
3º	R. Madrid
4º	Sevilla

Vemos como basándonos en el Linear Ordering Problem, la liga la habría ganado el Atlético de Madrid, quedando el Barcelona 2º y el 3º y 4º puesto no variaría respecto a la disposición normal.

Otra forma análoga de interpretar la matriz y obtener un resultado de ordenación sería la siguiente:

$$\begin{array}{cccc} B & M & AT & S \\ \left(\begin{array}{cccc} - & 5 & 1 & 5 \\ 4 & - & 2 & 5 \\ 2 & 2 & - & 4 \\ 2 & 4 & 4 & - \end{array}\right) & & & \\ 8 & 11 & 7 & 14 \end{array}$$

Sumamos por columnas, el resultado que obtenemos por ejemplo en la matriz observada anteriormente sería el siguiente: Barcelona 8, Real Madrid 11, Atlético de Madrid 7 y Sevilla 14. A la hora de ordenarlo seguimos el mismo criterio de que el mejor es el menor número, pues en este caso significa que le han metido menos goles. Con lo que el mejor sería el Atlético de Madrid, seguido del Barcelona, Real Madrid, y por último el Sevilla, justo la misma clasificación obtenida anteriormente.

Ahora vamos a hablar un poco más del Ordenamiento Lineal, vamos a ver más interpretaciones y la formulación.

Primeramente vamos a definirlo de una manera más técnica:

Sea $D_n = (V_n, A_n)$ el total de diagramas contenidos en n nodos, es decir, el gráfico dirigido con el conjunto de nodos $V_n = \{1,2, ..., n\}$ y para cada par de nodos *i* y *j*, hay un arco (*i*, *j*) de *i* a *j* y un arco (*j*, *i*) de *j* a *i*.

Un torneo *T* en A_n consiste en un subconjunto de arcos que contiene para cada par de nodos *i* y *j*, o un arco (*i*, *j*) o uno (*j*, *i*), pero no ambos.

Un torneo acíclico es un torneo sin ciclos dirigidos, es decir, que no contiene un conjunto de arco de la forma $\{(v_1, v_2), (v_2, v_3), ..., (v_k, v_1)\}$ para $k > 1$ y nodos distintos $v_1, v_2, ..., v_k$.

Un ordenamiento lineal de los nodos $\{1,2, ..., n\}$ es una ordenación de los nodos dados como secuencia lineal, o equivalentemente, como una permutación de los nodos.

Denotamos el ordenamiento lineal que clasifica el nodo *v1* primero, *v2* segundo, etc., y v_n último por $v_1, v_2 ..., v_n$ y escribir $v_i \prec v_j$ si el nodo vi se clasifica antes del nodo v_j.

Si σ denota una relación lineal ordenando entonces σ *(i)*, da la posición del nodo *i* en este ordenamiento.

Es fácil ver que un torneo acíclico *T* , *An* corresponde a un ordenamiento lineal de los nodos de V_n y viceversa: el nodo clasificado primero es el que no introduce arcos en *T*, el nodo clasificado en segundo lugar es el que tiene un arco de entrada (Es decir, desde el nodo primero), etc., y el nodo clasificado en último lugar es el que no sale arcos en *T*.

Por lo general, las relaciones de orden se ponderan y tenemos ponderaciones *Cij* dando el beneficio o coste resultante cuando el nodo *i* se clasifica antes del nodo *j* o, de forma equivalente, cuando el arco (*i, j*) está contenido en el torneo acíclico.

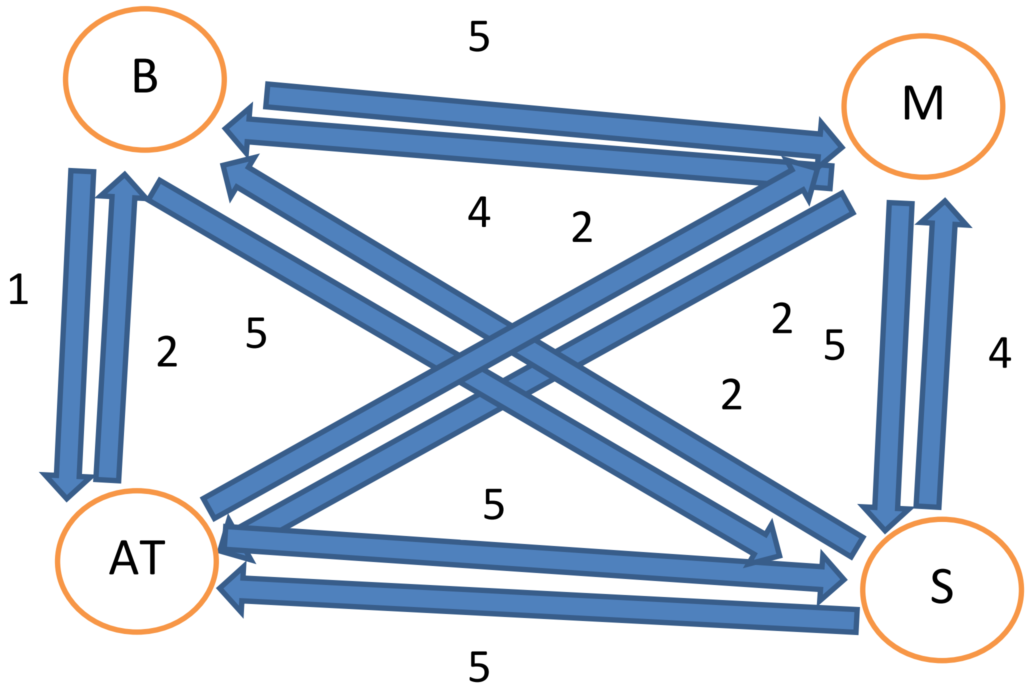
El problema de ordenamiento lineal se define:

Dado el grafo dirigido completo Dn = (Vn, An) con pesos de arco *Cij*, para cada par i, j ∈ Vn, calcula un torneo acíclico (que cuando se llega al nodo final se termina, no vuelve al principio) de expansión T en An tal que Σ (i, j) ∈ T Cij sea lo más grande posible.

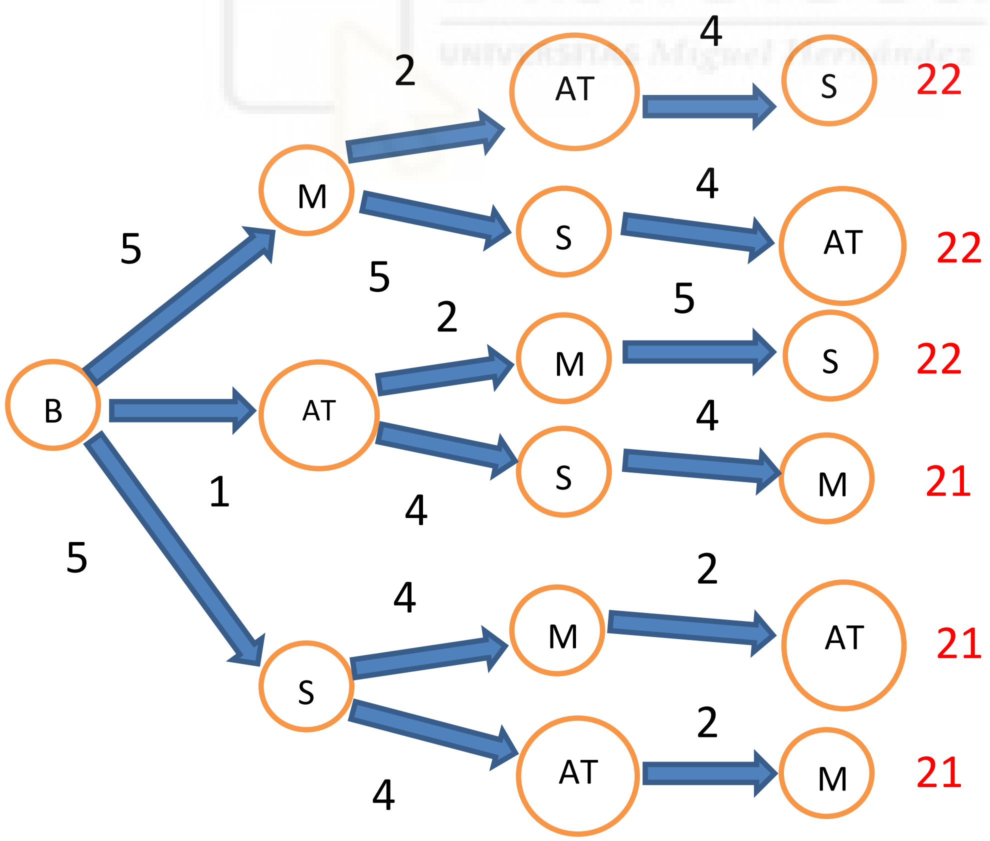
Aquí vamos a ver un ejemplo de cómo se plantearía y se resolvería el problema con grafos. Vamos a seguir con el ejemplo del fútbol, y vamos a partir de la matriz inicial:

$$
\begin{array}{c} \\ B \\ M \\ AT \\ S \end{array}
\begin{array}{c}
\begin{array}{cccc} B & M & AT & S \end{array} \\
\begin{pmatrix}
- & 5 & 1 & 5 \\
4 & - & 2 & 5 \\
2 & 2 & - & 4 \\
2 & 4 & 4 & -
\end{pmatrix}
\end{array}
$$

El grafo dirigido se nos quedaría de la siguiente manera:

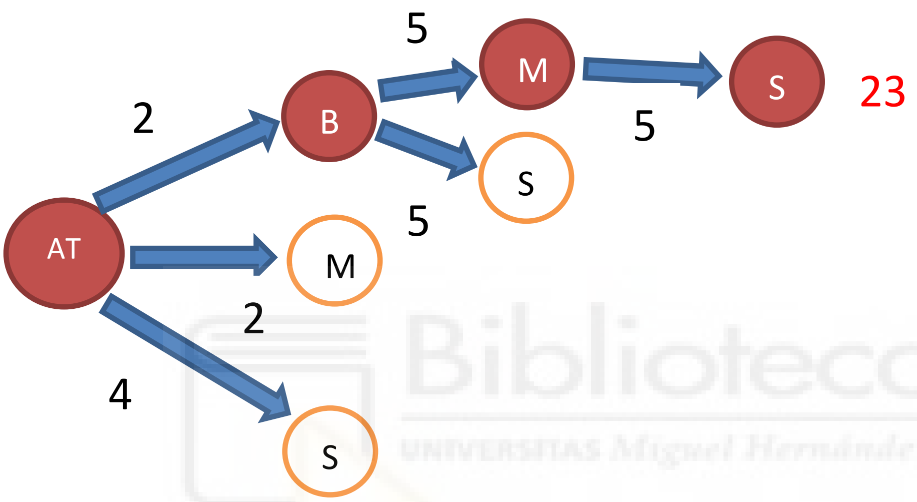


El torneo acíclico quedaría de la siguiente manera:



Si observamos el torneo, tenemos que quedarnos con el número que está en rojo a la derecha de los nodos finales, en concreto nos interesa que sea la mayor posible, por ejemplo el 22 primero sale de sumar 5+1+5+2+5+4.

Si hiciéramos esto con las 23 restantes matrices, obtendríamos 23 grafos y 23 torneos diferentes. Al final el máximo número al final de cada nodo daría 23, que coincidiría con que el Atlético de Madrid es el primero, el segundo es el Barcelona, el tercero el Real Madrid y por último el Sevilla, se quedaría como se muestra en el siguiente torneo:



Son dos maneras diferentes de resolver el problema de Ordenamiento Óptimo.

Para entender aún mejor el problema de Ordenamiento Óptimo, vamos a dar una definición en términos de optimización lineal:

Se define:

$$X_{ij} \begin{cases} 1 & si\ i\ antes\ que\ j \\ 0 & en\ caso\ contrario \end{cases}$$

Y entonces, el problema del LOP sería:

$$\max \sum_{(i,j)\in An} CijXij$$

s.a.

$$Xij + Xji = 1, \forall\, i, j\ \in Vn, i < j$$

$$Xij + Xjk + Xki \leq 2, \forall\, i, j, k\ \in Vn, i < j, i < k, j \neq k$$

$$Xij \in \{0,1\} \forall \, i,j \in Vn$$

Ahora ya podemos plantear y resolver el problema, es decir trasladamos el anterior planteamiento general a nuestro ejemplo de fútbol, y lo resolveremos con Matlab, el planteamiento del problema quedaría como sigue:

MAX
$5_{BM}+X_{BAT}+5X_{BS}+4X_{MB}+2X_{MAT}+5X_{MS}+2X_{ATB}+2X_{ATM}+4X_{ATS}+2X_{SB}+4X_{SM}+4X_{SAT}$

s.a.

$X_{BM}+X_{MB}=1$

$X_{BAT}+X_{ATB}=1$

$X_{BS}+X_{SB}=1$

$X_{MAT}+X_{ATM}=1$

$X_{MS}+X_{SM}=1$

$X_{ATS}+X_{SAT}=1$

$X_{BM}+X_{MAT}+X_{ATM}\leq 2$

$X_{BM}+X_{MS}+X_{SM}\leq 2$

$X_{BAT}+X_{ATM}+X_{MAT}\leq 2$

$X_{BAT}+X_{ATS}+X_{SAT}\leq 2$

$X_{BS}+X_{SAT}+X_{ATS}\leq 2$

$X_{BS}+X_{SM}+X_{MS}\leq 2$

$X_{MAT}+X_{ATS}+X_{SAT}\leq 2$

$X_{MS}+X_{SAT}+X_{ATS}\leq 2$

Con la versión actual de Matlab la función que usamos para incluir variables enteras es “intlinprog”. En realidad esta función admite variables continuas y enteras, por lo que hay que declarar cuáles son las variables enteras (es lo que se llama intcon), en este caso todas las variables son enteras.

Además para que Matlab las considere binarias (con valor 0 ó 1), tendremos que poner cota inferior 0 a esas variables y cota superior 1.

```
%Linear Ordering Problem
f=[-5;-1;-5;-4;-2;-5;-2;-2;-4;-2;-4;-4]
intcon=1:12;
A=[1,0,0,0,1,0,0,1,0,0,0,0;
   1,0,0,0,0,1,0,0,0,0,1,0;
   0,1,0,0,1,0,0,1,0,0,0,0;
   0,1,0,0,0,0,0,0,1,0,0,1;
   0,0,1,0,0,0,0,0,1,0,0,1;
   0,0,1,0,0,1,0,0,0,0,1,0;
   0,0,0,0,1,0,0,0,1,0,0,1;
   0,0,0,0,0,1,0,0,1,0,0,1];
b=[2 2 2 2 2 2 2 2];
Aeq=[1,0,0,1,0,0,0,0,0,0,0,0;
     0,1,0,0,0,0,1,0,0,0,0,0;
     0,0,1,0,0,0,0,0,0,1,0,0;
     0,0,0,0,1,0,0,1,0,0,0,0;
     0,0,0,0,0,1,0,0,0,0,1,0;
     0,0,0,0,0,0,0,0,1,0,0,1];
beq=[1 1 1 1 1 1];
lb=zeros(1,12);
ub=ones(1,12);
[x,fval]=intlinprog(f,intcon,A,b,Aeq,beq,lb,ub)
```

El resultado que obtenemos es el siguiente:

```
x =

     1
     0
     1
     0
     0
     1
     1
     1
     0
     0
     0
     1


fval =

   -23
```

Es decir las variables que toman el valor 1 son:

X_{BM}, X_{BS}, X_{MS}, X_{ATB}, X_{ATM} y X_{SAT}

Vemos como obtenemos el mismo resultado mencionado anteriormente, donde las suma de los elementos de la diagonal principal da 23 (el máximo), y la clasificación quedaría como hemos comentado antes.

Para concluir el Ordenamiento Óptimo, vamos a comentar varios *"papers"* (artículos) relacionados con el tema.

Hemos considerado interesante un ejemplo de a que se aplica el Ordenamiento Óptimo (Linear Ordering Problem) obtenido del libro:

- *The Linear Ordering Problem: Methods in Combinatorial Optimization (Applied Mathematical Sciences 175) del autor S.S. Antman J.E. Marsden L. Sirovich. (2011).* En concreto en la *página 7* se encuentra el mencionado ejemplo y trata sobre la percepción de la corrupción.

La organización Internacional de Transparencia publica un índice anual de percepción de la corrupción que clasifica a más de 150 países por su nivel percibido de corrupción.

Este índice se calcula a partir de evaluaciones de expertos y encuestas de opinión.

Las respectivas evaluaciones y encuestas sólo consideran un subconjunto de todos los países y, por lo tanto, también pueden considerarse como un orden parcial.

En el problema de ordenamiento lineal se utiliza para agregar estos rankings parciales. Se muestra que la solución de la el problema de ordenación concuerda con el ranking según el índice en gran medida, pero muestra diferencias interesantes para algunos países.

- *Campos, V., Glover, F. Laguna, M. and Martí, R.: An experimental evaluation of a scatter search for the linear ordering problem, Journal of Global Optimizationz 21 (2001), 397–414.*

En este artículo se habla sobre metodologías de búsqueda para implementar nuevas maneras para aplicar el problema de Ordenamiento Lineal (Linear

Ordering Problem, o LOP para simplificar). El problema de ordenamiento lineal (LOP) ha generado una cantidad considerable de interés de investigación a lo largo de los años, como se documenta en Grotschel, et al. (1984) y Chanas y Kobylanski (1996). Debido a su relevancia práctica y teórica para una amplia gama de aplicaciones de optimización global, los autores utilizan este problema como un caso de prueba para sus estrategias y mecanismos de búsqueda.

En economía, el problema de Ordenamiento Lineal es equivalente al llamado problema de triangulación para las tablas de entrada-salida, que se puede describir como sigue. La economía de una región (generalmente un país) se divide en m sectores y se construye una tabla de insumo-producto m × m E donde la entrada eij denota la cantidad de entregas (en valor monetario) del sector i al sector j en un determinado año. El problema de triangulación consiste entonces en permutar simultáneamente las filas y columnas de E, para hacer la suma de las entradas por encima de la diagonal principal lo más grande posible. Una solución óptima ordena a los sectores de tal manera que los proveedores (es decir, los sectores que tienden a producir materiales para otras industrias) vienen primero, seguidos por los consumidores (es decir, los sectores que tienden a ser las industrias del producto final que entregan su producción Principalmente a usuarios finales).

El artículo concluye diciendo que el objetivo de esta investigación ha sido avanzar en el estado actual del conocimiento sobre las implementaciones de los procedimientos de búsqueda de dispersión para la optimización global. A diferencia de otras investigaciones, la búsqueda de dispersión no ha sido ampliamente estudiada. Han logrado determinar una forma de diversificación que resulta bastante efectiva en lo referente a este escenario.

- *García, C.G., Pérez-Brito, D., Campos, V. and Martí, R.: Variable neighborhood search for the linear ordering problem, Computers and Operations Research 33 (2006),3549–3565*

Este artículo habla que el problema de Ordenamiento Lineal ha generado una cantidad considerable de interés en la investigación desde 1958, cuando

Chenery y Watanabe, en 1958 delinearon algunas ideas sobre cómo obtener soluciones para este problema.

Becker, en 1967, propone una heurística basada en el cálculo de cocientes para clasificar cada sector de la economía, el sector primario (agricultura, ganadería, pesca, minería y producción energética), el sector secundario (industria, construcción y manufactura), y el sector terciario (comercio, bancos, educación, cultura y servicios persona a persona). El sector con mayor cociente es el más alto. La columna y la fila correspondientes se suprimen de la matriz y el procedimiento se aplica a los sectores restantes. El método de Becker es bastante rápido y produce resultados razonables considerando su simplicidad.

Grotschel en 1984, describe un algoritmo de solución exacta para el problema de Ordenamiento Lineal, que podría ser considerado como el primer verdadero algoritmo de ramificación y corte. Este método explora la descripción lineal del problema de Ordenamiento Lineal y es capaz de aplicarse y obtener soluciones óptimas para un conjunto de situaciones del mundo real. Reinelt, en 1985, resume los métodos y aplicaciones relevantes a este problema. Este autor también mantiene una biblioteca de dominio público, llamada LOLIB desde 1997, que cuenta con 49 posibles situaciones de tablas input-output de sectores de la economía europea. Este sitio web también contiene soluciones óptimas para estas situaciones. Utilizaremos estos resultados en nuestros experimentos computacionales para medir la calidad de los métodos heurísticos. Ha habido otras contribuciones importantes en el contexto de los métodos exactos para el problema de Ordenamiento Lineal. Por ejemplo, Mitchell y Borchers, en el año 2000, proponen un algoritmo basado en un método de punto interior primal-dual y método simplex para resolver el problema. Sus resultados computacionales muestran la eficacia del procedimiento propuesto. Sin embargo en la investigación que llevan a cabo en este artículo se centran en desarrollar un método heurístico para este problema.

- *Charon, I. and Hudry, O.: A survey on the linear ordering problem for weighted or unweighted tournaments, 4OR 5 (2007), 5–60.*

Este artículo nos habla de dos problemas principalmente, el problema de Kemeny y el problema de Slater. Estos tienen un mismo modelo gráfico teórico, aunque sus orígenes y significados son diferentes.

Ambos se ocupan de órdenes lineales, es decir, relaciones binarias transitivas asimétricas completas.

El primero, el problema de Kemeny (Kemeny, 1959), también se atribuye a Condorcet (1785), el cual en la introducción del trabajo ya hemos comentado su paradoja.

Resumiendo, en este artículo, nos habla de resultados, conjeturas y problemas abiertos que tratan los aspectos combinatorios y algorítmicos del problema de ordenamiento lineal.

Estos problemas consisten en encontrar un ordenamiento lineal que este a una distancia mínima de un torneo. Nos muestran cómo se pueden utilizar para modelar un problema de agregación, el cual consiste en pasar de preferencias individuales definidas en un conjunto de candidatos a una clasificación final de estos candidatos.

4. CADENA DE MARKOV

Vamos a ver cómo quedaría ahora este mismo ejemplo usando una Cadena de Markov. Vamos a usar nuestra primera matriz, (que será nuestra matriz de transición), el problema es que para poder usar Cadena de Markov la suma de las filas de la matriz ha de ser igual a 1, luego necesitamos transformarla. Lo haremos de la siguiente forma:

$$\begin{pmatrix} - & 5 & 1 & 5 \\ 4 & - & 2 & 5 \\ 2 & 2 & - & 4 \\ 2 & 4 & 4 & - \end{pmatrix} \Longrightarrow \begin{pmatrix} - & 6/22 & 10/22 & 6/22 \\ 7/22 & - & 9/22 & 6/22 \\ 6/16 & 6/16 & - & 4/16 \\ 8/20 & 6/20 & 6/20 & - \end{pmatrix}$$

Lo que hemos hecho es sumar la fila y a este número restar cada elemento de la matriz, y este sería el numerador del elemento de la nueva matriz y para el denominador, cogemos la suma de los elementos de la fila de la nueva matriz.

Por ejemplo el elemento a_{12} de la nueva matriz saldría de la siguiente manera: Sumamos la primera fila, 5 + 1 +5 que es igual a 11. A este 11 le restamos el elemento a_{12} de la matriz que antigua y obtenemos el 6, que será el numerador del elemento a_{12} de la nueva matriz. Hacemos lo mismo con el elemento a_{13} y a_{14}, y después sumamos la fila de la nueva matriz, el número que obtenemos es el nuevo denominador de esta fila. Lo hacemos porque además de cómo hemos mencionado anteriormente las filas deben de sumar uno, en Cadena de Markov, a diferencia del Linear Ordering Problem, para ordenar a los equipos, el mejor es el mayor número y no el menor.

Simplificando un poco la matriz de transición se nos queda:

$$T= \begin{array}{c} \\ \text{B} \\ \text{M} \\ \text{AT} \\ \text{S} \end{array} \begin{array}{c} \begin{array}{cccc} \text{B} & \text{M} & \text{AT} & \text{S} \end{array} \\ \begin{pmatrix} - & 3/11 & 5/11 & 3/11 \\ 7/22 & - & 9/22 & 3/11 \\ 3/8 & 3/8 & - & 1/4 \\ 2/5 & 3/10 & 3/10 & - \end{pmatrix} \end{array}$$

Analizamos un poco la matriz. Cada equipo (cliente) tiene su opinión sobre la clasificación, vamos por filas. Por ejemplo, para el Barcelona el mejor sería el Atlético (probabilidad más alta), y estarían empatados el Sevilla y el Madrid, el Madrid (segunda fila) opinaría lo mismo. En la tercera fila, el Atlético ya discrepa, ya que para él, Barça y Madrid son iguales, y el peor es el Sevilla. En la última fila, el Sevilla opina que le mejor es el Barcelona, y Madrid y Atlético son iguales.

Antes de formular el problema para pasar a resolverlo con el RStudio, vamos a definir una Cadena de Markov.

Una Cadena de Markov es un proceso estocástico con espacio de estado discreto, (ejemplos de estados discretos serían el número de llamadas al 091 en el momento t o el número de piezas blancas en un tablero de ajedrez después de la t-ésima jugada) que cumple la propiedad de la pérdida de memoria de Markov. Es decir, la probabilidad del estado actual solo depende del estado anterior y no de toda la evolución.

Vemos un ejemplo de cadena de Markov: La probabilidad de que llueva mañana depende de la humedad de hoy, si dependiera de los últimos 5 días no (sé lo que pasa en el estado anterior).

Existen cadenas de Markov con tiempo discreto y continuo. Si el estado fuera la distancia no es una cadena de Markov.

Si la probabilidad de que $P(X_{n+1} = j|X_n = i) = P_{ij}, \forall n$, es decir, no depende de n, entonces,

- La cadena de Markov se dice homogénea. No depende de n.
- A las probabilidades P_{ij} se les dice probabilidades de transición. La matriz de estas probabilidades es la matriz de transición y en ella cada fila representa el estado actual y cada columna el estado futuro.
- La cadena se puede representar por un grafo que se llama diagrama de transición de estados.

- La cadena se puede representar por la matriz de transición. La dimensión de la matriz es el número de estados, y esta es siempre cuadrada.

$$\begin{pmatrix} P00 & \cdots & P0\,|E| \\ \vdots & \ddots & \vdots \\ P\,|E|0 & \cdots & P\,|E||E| \end{pmatrix}$$

La suma de cualquier fila de *P* es 1 y cualquier matriz que cumple que la suma de cada una de sus filas es 1 se dice estocástica. Además esta matriz no puede tener una columna con todo ceros. Si todos son ceros no se da ningún estado, no puede ser una matriz estocástica. Una columna si puede sumar más de 1.

Si un vector indica su estado inicial w, entonces wP es el estado siguiente.

Al igual que hemos hecho con el Ordenamiento Óptimo, vamos ahora a definir de una manera menos técnica lo que es una Cadena de Markov, para entenderlo mejor.

Explicaremos en primer lugar lo que es un proceso estocástico. Este consiste en una serie de eventos que se desenvuelven en el momento en el cual cualquier etapa contiene algún elemento que depende de la fortuna.

Ejemplos para entenderlo podría ser el precio de las acciones que cotizan en bolsa o las condiciones meteorológicas durante días consecutivos en un determinado lugar.

Lo que queremos decir es que el resultado en cualquier etapa no es independiente de los resultados anteriores, eso es un proceso estocástico, el precio de una acción de un día depende en parte de con qué precio cerro el día anterior, y el tiempo un día no es independiente de el del día de antes, sino que se ve afectado.

Ahora ya podemos definir una CADENA DE MARKOV, es un proceso estocástico, donde los resultados dependen de otros (el resultado de una etapa solo depende de la anterior).

Las cadenas de Markov tienen memoria, “recuerdan” el último evento, y eso condiciona los posteriores.

Una propiedad de las Cadenas de Markov, es que el estado t+1 solo depende del estado en t y no de la evolución anterior del sistema.

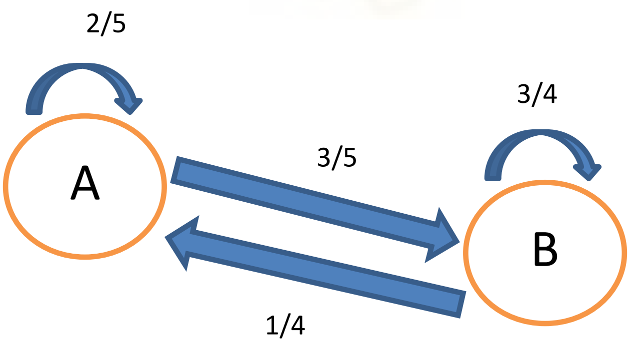
Ahora vamos a definiremos la matriz de transición, fundamental para trabajar con cadenas de Markov.

Para entenderlo, vamos a considerar unas elecciones políticas en un país, donde solo hay dos partidos, el A y el B.

Si A gana las elecciones, el país se encuentra en el estado A, y lo mismo ocurre si gana B.

- Si A es el que ha ganado y tiene el poder, existe una probabilidad de 2/5 de que gane las próximas y de 3/5 de que las pierda frente a B.
- Si B es el que tiene el poder, hay una probabilidad de 1/4 de que gane A las próximas, y una de 3/4 de que siga estando B en el poder.

La sucesión de elecciones forman una cadena de Markov:



- Los círculos se denominan nodos, y representan los estados del proceso.
- Las flechas son los arcos, y representan la probabilidad de cambiar de un estado a otro.

Esto, lo podemos representar en forma de matriz, la llamada matriz de transición, que sería:

$$\begin{array}{cc} & \text{Resultado próxima elección} \\ & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} \text{Resultado última elección} \end{array} \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} 2/5 & 3/5 \\ 1/4 & 3/4 \end{pmatrix} \end{array}$$

Cabe indicar que la suma de las filas debe ser igual a 1, pues cada uno de los elementos de una fila representa la probabilidad de que el sistema pase a uno de los estados, y además las probabilidades deben ser mayores o iguales que cero.

Por último, volviendo al ejemplo inicial sobre la liga de fútbol, vamos a formular y resolver el problema, y así poder comparar los resultados con los obtenidos mediante Ordenamiento Óptimo.

Con Cadena de Markov obtenemos el mismo "problema" que con Ordenamiento Óptimo, es decir, que no están deacuerdo todos, para resolverlo vamos a utilizar el RStudio con la librería markovchain.

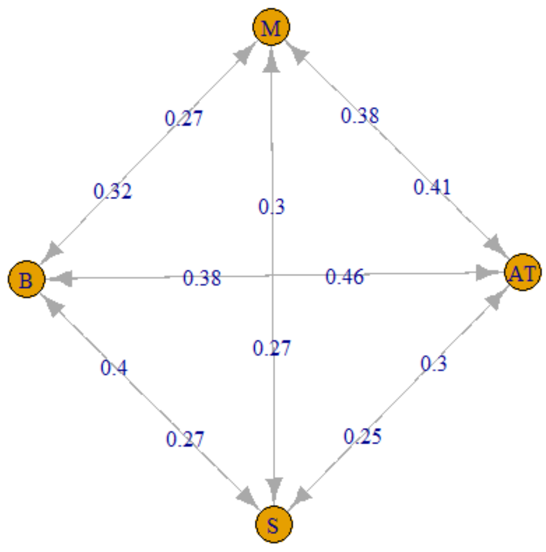
```
> install.packages("markovchain")

> library(markovchain)
> estados<-c("B","M","AT","S")
> matrizTransicion=matrix(data=c(0,0.27,0.46,0.27,0.32,0,0.41,0.27,0.375,0.375,0,0.25,0.4,0.3,0.3,0),
+                         byrow=TRUE,nrow=4,dimnames=list(estados,estados))
> CadenaMarkov=new("markovchain",states=estados,byrow=TRUE,transitionMatrix=matrizTransicion,name="Fútbol")
> show(CadenaMarkov)
Fútbol
 A  4 - dimensional discrete Markov Chain with following states
 B M AT S
 The transition matrix   (by rows)  is defined as follows
       B     M    AT    S
B  0.000 0.270 0.46 0.27
M  0.320 0.000 0.41 0.27
AT 0.375 0.375 0.00 0.25
S  0.400 0.300 0.30 0.00
```

Mediante el estado de equilibrio vemos la clasificación óptima:

```
> steadyStates(CadenaMarkov) #estado equilibrio
             B         M        AT         S
[1,] 0.2668716 0.2409957 0.2840069 0.2081259
```

El paquete markovchain del RStudio también nos proporciona el grafo, que quedaría como se muestra a continuación:



Con Cadena de Markov la clasificación final se nos quedaría:

Clasificación con CADENA DE MARKOV

PUESTO	EQUIPO
1º	Atlético de Madrid
2º	Barcelona
3º	R. Madrid
4º	Sevilla

Vemos que con Cadena de Markov obtenemos el mismo resultado que con Ordenamiento Óptimo. Aunque en general no será así.

Ya podemos comparar los resultados obtenidos por ambos métodos. En este ejemplo es indiferente elegir entre Ordenamiento Óptimo o Cadena de Markov, ya que al final obtenemos el mismo resultado.

Si hubiéramos obtenido resultados diferentes, de cara a elegir entre Cadena de Markov u Ordenamiento Óptimo, tendríamos que ver las ventajas y desventajas de cada método, y elegir uno en función de lo que más nos interese.

Continuando con Cadena de Markov, hemos considerado interesante comentar un ejemplo de cómo se aplica la Cadena de Markov para hacer rankings, se trata de la página web de Google, obtenido en el siguiente enlace:

http://www.fcfm.buap.mx/assets/docs/docencia/tesis/ma/MariaGuadalupeTorresGonzalez.pdf

Nunca te has preguntado, ¿por qué esa ordenación cuando introduces algo que te interesa en el buscador?, ¿por qué normalmente devuelve primero las más relevantes? Pues bien es un ranking que se elabora automáticamente basado en una Cadena de Markov.

En Google están registradas una serie de webs, que se actualizan periódicamente a través de un robot que busca e incorpora páginas. Podemos decir que el registro de Google posee N sitios (miles de millones).

Google asocia a cada página web registrada un número que va desde el 1 hasta el N, y se construye un grafo G con conjunto de vértices E={0,1,….,N} y aristas de acuerdo a lo siguiente:

- No hay aristas desde un nodo hacía ese mismo nodo.
- Hay una única arista desde 0 hasta cada $i \in \{1,\ldots,N\}$.
- Hay una única arista desde cada $i \in \{1,\ldots,N\}$ hasta 0.
- Para *i, j* positivos y distintos, hay una única arista de *i* a *j* si, y sólo si, en la página web *i* hay algún enlace referenciado a la página web *j*.

Se define *S(i)* como una unidad mayor que la cantidad de páginas web a las que puede accederse por enlace desde la página i, es decir $S(i) \geq 1$ para todo $i \in E$.

A continuación se define la matriz *P* con índices en *E* de la siguiente forma:

Para todo $i \in E$, $P_{ii} = 0$.

Para todo $i \in \{1,\ldots.,N\}$, $P_{0i} = 1/N$.

Para todo $i \in \{1,\ldots.,N\}$,

$$P_{io} = \begin{cases} 1 & si \ \ S(i) = 1 \\ 1 - p & si \ \ S(i) > 1 \end{cases}$$

Para todos $i,j \in \{1,\dots,N\}$ con $i \neq j$,

$$P_{ij} = \begin{cases} \frac{p}{S(i)-1} & si \ \ hay\ arista\ de\ i\ a\ j\ en\ G \\ 0 & en\ caso\ contrario \end{cases}$$

Google solo necesita construir la matriz P en base a los enlaces entre páginas, esta matriz es estocástica, y corresponde a una cadena de Markov irreducible y aperiódica que tiene todos sus estados persistentes positivos, y solo admite una sola distribución estacionaria v.

Google interpreta v_j como el ranking de la página *j*, supone que cuanto más relevante es la página, más visitas tiene a lo largo del recorrido por el grafo *G*, midiendo las frecuencias de cada visita por v_j. Por lo tanto, el orden viene dado por el valor en el estado de equilibrio.

Cabe destacar que la distribución estacionaria v se encuentra calculando P^n para *n* suficientemente grande y observamos su primera fila. Además la construcción de v se efectúa periódicamente, después de dar de alta a nuevas páginas en los registros y de actualizar la matriz *P*, entonces el *ranking* ya estaría disponible al ordenar una búsqueda.

Al igual que en la sección anterior, para finalizar con Cadena de Markov, vamos a comentar otros artículos interesantes relacionados con el tema.

http://e-spacio.uned.es/fez/eserv/bibliuned:20379/cadenas_markov.pdf

- *GONZALEZ CARMONA, ANDRES, SAENZ BARRIO OSCAR: CADENAS DE MARKOV APLICADAS AL DIAGNÓSTICO Y TRATAMIENTO DE CONDUCTAS Y PROCESOS DE APRENDIZAJE, 32-45*

Este artículo habla sobre que una serie de posibilidades que ofrecen las cadenas de Markov, las cuales permiten eliminar problemas y realizar un análisis de predicción.

Comenta que las matrices de transición se han revelado como unos instrumentos muy valiosos en el estudio de la conducta, permiten predecir, mediante cadenas probabilísticas, la transición de una fase a otra dentro de una secuencia, y anticipar el estado más probable en el cual un sujeto va a ser observado cuando tal proceso alcanza su equilibrio.

Los autores han analizado matrices de datos, los cuales han sido probados por ellos mismos en secuencias de comportamiento bien definidas como: agresividad (Francés, González y Sáenz, 1987), dependencia (Francés y Sáenz, 1987), competencia (Francés y Sáenz, 1988), resolución de problemas (Alonso, González y Sáenz, 1988). Creen que ha llegado el momento de generalizar un modelo diagnóstico que parece responder a la respuesta de la idea del matemático Verstappen expuesta en la V Reunión del ICME (1984). Verstappen afirmaba que el objetivo supremo para los profesores era identificar alguna suerte de estructura que permitiera apoyar cambios en la acción pedagógica, de tal forma que se pudiera penetrar en los perspectivas de los estudiantes, y estimular al profesor como investigador en el aula (Carss, 1986).

Nos comentan que primero categorizan la conducta, por ejemplo, la agresividad la han estudiado en tres situaciones escolares: controlada, semicontrolada y libre. Posteriormente se recogen los datos mediante la acotación de conductas observadas por unidad de tiempo, pero aparte también se requiere que los datos estén ordenados en relación a la aparición de cada conducta.

Si en vez de codificar las observaciones en tablas de frecuencia se codifican en forma matricial, su información es mucho más rica. Las matrices de frecuencias de transición *(n¡j)* vienen definidas de acuerdo con la condición de que cada elemento *ni,j* sea el número de transiciones observadas desde el estado *Ei* al *Ej*: *nij = (Ei, Ej).*

Se pueden conseguir la siguientes tipos de matrices, de perfil individual (consiste en la suma de todas aquellas matrices que codifican un determinado tipo de comportamiento), matriz de conducta-tipo (comparar la misma secuencia o conducta en diferentes situaciones o estados) y la matriz de conducta generalizada (El cálculo matricial permite desvelar la cadena comportamental derivada de sujetos X conductas x situaciones).

Algunas conclusiones del artículo son que las matrices de transición constituyen el tratamiento de elección cuando se quiere estudiar la conducta, como una sucesión en la que un estado sustituye a otro o a sí mismo.

Su expresión en términos de probabilidad confiere a las matrices de transición un valor pronóstico de la conducta de un individuo o un grupo, por cuanto en función de su conducta anterior se puede predecir con qué probabilidad pasará a un estado a otro.

Se ha comprobado que la teoría de cadenas de Markov permite calcular en determinados casos el vector límite de un proceso secuencializado; conlleva que cualquiera que sea la conducta de la que parte un sujeto o un grupo, llegará a la situación del vector límite con una probabilidad dada, una vez alcanzado el estado de equilibrio.

- *UTILIZAR LAS CADENAS DE MARKOV PARA TOMAR MEJORES DECISIONES. Tirado, Maribel. 22/01/2014 (Artículo original de Michael Walker en Data Science Association).*

http://www.bigdatahispano.org/noticias/cadenas-de-markov-para-tomar-mejores-decisiones/

Un “paper” o artículo interesante sobre Cadenas de Markov, lo he encontrado en una web de “big data”. El artículo hace referencia a cómo podemos tomar decisiones óptimas, y como entran ahí las Cadenas de Markov para facilitar esta toma de decisiones.

Las Cadenas de Markov son apropiadas para crear e insertar procesos de toma de decisiones con diferentes alternativas y predecir así con una mayor exactitud posibles comportamientos que se den el futuro.

El trabajo del matemático Andrey Markov derivo en la Cadena de Markov, que no es más que una secuencia de valores de una variable aleatoria, en la que el valor de esa variable en el futuro se puede predecir basándose en datos actuales, datos pasados y probables condiciones futuras.

Las Cadenas de Markov son interesantes para la toma de decisiones, sobretodo en procesos donde intervienen la suerte y las decisiones previas.

Podríamos considerar las Cadenas de Markov como un conjunto de fórmulas, con un número y un peso asignado para cada una de ellas. El objetivo es encontrar la distribución estacionaria, y a continuación obtener la probabilidad de que tenga lugar una fórmula A dada una fórmula B que es cierta.

Ejemplos de aplicaciones de las Cadenas de Markov son:

- *Investigación en ciencias de la salud:* Desde comienzos de los 80, el SIDA se ha convertido en una de las mayores pandemias de nuestra época. Nos muestran cómo podemos aplicar una Cadena de Markov a un análisis de la epidemia de VIH-SIDA, definiendo los estados del proceso, cada sujeto de la población puede estar exclusivamente en uno de los siguientes estados: susceptible (S), infectado por VIH (VIH), con SIDA(SIDA) o muerto como consecuencia de la enfermedad (M). El sujeto susceptible no está infectado (estado 1, -S-). Tras el contacto con un infectado desarrollará la infección y podrá contagiar a otros individuos de la población (estado 2, -VIH-). Cuando el sujeto infectado desarrolla la enfermedad se convertirá en un caso de SIDA (estado 3, -SIDA-), pudiendo morir por causa de la enfermedad (estado 4, -M-). Los estados S, VIH y SIDA son transitorios, puesto que el individuo nunca volverá a ellos una vez que los abandone. El estado M en cambio, es absorbente, ya que los sujetos que lo alcanzan permanecen en el sin posibilidad de cambiar a otro.

- *Predicciones meteorológicas*: Poder conocer las condiciones meteorológicas siempre ha sido algo interés común, ya que nos permite hacer planes a corto/medio plazo. Podemos definir los estados como soleado y nublado por ejemplo. Si sabemos que la probabilidad de que haga sol el miércoles es 0.2, y la probabilidad de que esté nublado es 0.8, podemos calcular si estará nublado el jueves, viernes y sábado.
- *Análisis de empresas*: Una manera diferente de controlar ciertos factores de la gestión de un negocio es realizar aproximaciones o previsiones en base a la utilización de cadenas de Markov. Es útil para previsiones a largo o muy largo plazo. A diferencia del método clásico de utilizar el año inmediatamente anterior como guía, la cadena de Markov utiliza todos los estados anteriores para determinar una evolución más realista de lo que cabe esperar de los próximos ejercicios. Dado el gran número y variedad de cifras existentes en los balances lo mejor es realizar el estudio en base a los ratios. Esto descubre más posibilidades pues también podemos utilizar los ratios de RRHH, que cuantifican valores cualitativos y miden factores humanos.
- *Predicción de pérdida de clientes:* Por ejemplo en un país como Colombia existen tres operadores principales de telefonía móvil como lo son tigo, Comcel y movistar (estos tres serían los tres estados). Conociendo los porcentajes actuales que tiene cada operador en el mercado actual, y sabiendo las probabilidades de que tiene un cliente a día de hoy de permanecer en un estado (no cambiarse de compañía), o pasar a un estado diferente (cambiarse de compañía), podemos predecir si en el futuro este cliente se cambiara de compañía o no.

Otros ejemplos de casos donde se pueden aplicar las Cadenas de Markov son: algoritmos de posicionamiento web (la Cadena de Markov en este caso actúa para hacer predicciones futuras de navegación); en la bolsa en lo referente al tema de las acciones para optimizar la compra venta de estas; aplicarse a deportes como por ejemplo el béisbol; señal, procesamiento y reconocimiento de imágenes; reconocimiento de patrones; seguros de vida y diversas aplicaciones en marketing.

5. ENJAEZAMIENTO DE LOS CABALLOS DEL VINO

A continuación, vamos a aplicar los dos métodos vistos anteriormente de Ordenamiento Óptimo y Cadena de Markov al Concurso de Enjaezamiento de los Caballos del Vino de Caravaca de la Cruz (Murcia).

Pero en primer lugar, vamos a explicar un poco en que consiste la fiesta de los Caballos del Vino.

La información sobre la Fiesta esta extraída del siguiente enlace:

http://www.sitiosdeespana.es/Evento/caballos-del-vino#.WRsWqWjyhPY

Hoy en día son unas 60 peñas con sus respectivos caballos los que son participes de esta festividad. Cada uno de los caballos va acompañado de 4 mozos, los Caballos del Vino recorren las calles de Caravaca, y al final hacen una carrera cuando llegan a la subida al castillo.

La peculiaridad de esta fiesta es el enjaezamiento de los Caballos del Vino, realizadas con muchas piezas bordadas con seda, piedras y canutillo de oro y plata, estas se renuevan cada año y además están entalladas para cada caballo, con el fin que luzcan perfectas en ellos.

Haciendo un poco de historia, en sus orígenes, los caballos eran los que llevaban el vino hasta la capilla de la Vera Cruz (que esta tras la subida al castillo), este vino se usaba en la misa para bendecir las flores, y después se repartía entre los asistentes junto con estas. Con el paso de los años dejaron de utilizar los caballos para transportar el vino, pero la fiesta de los Caballos del Vino empezó a desarrollarse hasta el punto de que los caballos son los únicos protagonistas de la festividad.

Con el paso del tiempo, los Caballos del Vino han ido evolucionando, hasta hoy en día donde la fiesta es un triple concurso que consiste en valorar al caballo primeramente a pelo, es decir se valora la figura y porte de este; en segundo lugar se valora el enjaezamiento, es decir la belleza y calidad de cómo van vestidos, y por último la carrera, donde lo más importante es lo rápido y preciso que sea el caballo. Todo esto claro siguiendo unas reglas establecidas desde hace años.

Las fiestas de esta localidad de Murcia tienen lugar los días 1,2 y 3 de mayo, pero durante todo el año se le da "bombo" con múltiples actos que están relacionados.

La fiesta comienza el 1 de mayo con el concurso de "Caballo a pelo". Las peñas recorren las calles de la ciudad, y se premia al caballo en estado puro.

El día 2 es el día grande de la Fiesta de los Caballos del Vino. Comienza en la madrugada con el ritual de "vestir al caballo", donde se procede a poner las prendas al caballo con mucha cautela, de tal forma que las piezas le encajen perfectamente. Durante el día se pasean por las calles desfilando, y a mediodía, en la subida al castillo, tiene lugar la carrera de los Caballos del Vino. En la prueba se mide el tiempo que el caballo tarda en subir, acompañado de cuatro mozos que corren a su lado, agarrados a él, dos se colocan delante, y dos se colocan detrás. Si alguno de los mozos se suelta antes de cruzar la meta del caballo, se produce la eliminación de este. La subida son 80 metros de una pendiente pronunciada. Normalmente los caballos más veloces la hacen en unos ocho segundos. Otra de las dificultades de la prueba, es que la subida se encuentra repleta de público, este abre paso a los caballos y mozos cuando pasan y después se vuelve a cerrar (como si fuese un fuelle).

El último día, el 3 de mayo de mañana, es para los más pequeños, con pasacalles y carrera.

La Fiesta de los Caballos del Vino es candidata a ser declarada Patrimonio Cultural Inmaterial de la Humanidad por la UNESCO. Cada año la ciudad recibe cerca de 200.000 visitantes de todo el mundo con el fin de vivir una fiesta única y difícil de describir.

Una vez que conocemos un poco en que consiste la Fiesta, vamos a pasar a aplicar los métodos para la unificación de votaciones vistos a las votaciones de los Caballos del Vino, para ver cuál sería el orden óptimo y así poder saber el caballo ganador.

Vamos a partir de un grupo de 11 peñas, que corresponde a las votaciones de cada caballo, cada peña tiene que votar a los demás caballos (darles una posición), no pudiendo dársela a su propio caballo por razones obvias.

VOTACIONES DE LOS CABALLOS DEL VINO

Nº CABALLO	1	2	3	4	5	6	7	8	9	10	11
1	-	6	4	9	8	3	5	10	1	2	7
2	5	-	2	7	8	6	4	10	3	1	9
3	7	4	-	9	10	6	3	8	2	1	5
4	7	3	4	-	8	5	6	10	2	1	9
5	7	4	3	9	-	6	5	10	2	1	8
6	4	7	3	8	9	-	5	10	1	2	6
7	7	5	1	9	10	6	-	8	2	3	4
8	9	5	1	8	10	7	4	-	2	3	6
9	4	5	3	10	8	2	6	9	-	1	7
10	5	2	1	7	8	4	6	10	3	-	9
11	7	5	1	8	10	6	3	9	2	4	-

El número de caballo se corresponde con el siguiente nombre:

1. AMBICIOSO
2. CAMPEON
3. CAPRICHOSO
4. ESPARTACO
5. JUPITER
6. MEL-AZULES
7. SANGRINO
8. SOLTERON-TRIANA
9. TERRY
10. UNIVERSO
11. ZAMBRA

Los números de la tabla van del 1 al 11, siendo 1 la mejor posición posible, y 10 la peor (no puede ser 11, porque una peña no puede votar a su propio caballo).

Vamos a explicar un poco la tabla para entender bien la información que nos está proporcionando. Decir que quien vota es la peña a la que pertenece el caballo, posicionan a los demás caballos (de sus respectivas peñas), no pudiendo repetir una posición, es decir cada peña hace una ordenación o ranking de los caballos.

Cada peña a la que pertenece el caballo es como si fuese un cliente, y tiene su opinión sobre cuál es el mejor caballo (ordenamiento óptimo). En la primera fila, la peña a la que pertenece el caballo número 1, AMBICIOSO “dice” que el mejor es el número 9, TERRY (porque es al que coloca en la posición 1), y que el peor sería el número 8, SOLTERON-TRIANA (que lo coloca en la última posición, la 10). En la segunda fila, la peña a la que pertenece el caballo número 2, CAMPEON, hace su ordenamiento, para ellos el mejor caballo es el número 10, UNIVERSO (lo coloca en la posición 1), y en último lugar estaría el caballo número 8, SOLTERON-TRIANA (posición 10). Y así con el resto de filas. Al final cada peña a la que pertenece cada uno de los caballos tiene una ordenación sobre los caballos, y como no están de acuerdo, tenemos un problema.

Vamos a resolverlo por los dos métodos estudiados, primeramente como en ambos métodos partimos de una matriz, vamos a poner la tabla de los Caballos del Vino en forma de matriz.

$$\begin{pmatrix} 0 & 6 & 4 & 9 & 8 & 3 & 5 & 10 & 1 & 2 & 7 \\ 5 & 0 & 2 & 7 & 8 & 6 & 4 & 10 & 3 & 1 & 9 \\ 7 & 4 & 0 & 9 & 10 & 6 & 3 & 8 & 2 & 1 & 5 \\ 7 & 3 & 4 & 0 & 8 & 5 & 6 & 10 & 2 & 1 & 9 \\ 7 & 4 & 3 & 9 & 0 & 6 & 5 & 10 & 2 & 1 & 8 \\ 4 & 7 & 3 & 8 & 9 & 0 & 5 & 10 & 1 & 2 & 6 \\ 7 & 5 & 1 & 9 & 10 & 6 & 0 & 8 & 2 & 3 & 4 \\ 9 & 5 & 1 & 8 & 10 & 7 & 4 & 0 & 2 & 3 & 6 \\ 4 & 5 & 3 & 10 & 8 & 2 & 6 & 9 & 0 & 1 & 7 \\ 5 & 2 & 1 & 7 & 8 & 4 & 6 & 10 & 3 & 0 & 9 \\ 7 & 5 & 1 & 8 & 10 & 6 & 3 & 9 & 2 & 4 & 0 \end{pmatrix}$$

Ya tenemos la matriz inicial de la que partimos, vamos a pasar a plantear y resolver el problema que nos permita conocer cuál es el orden óptimo de los Caballos del Vino (ranking).

Primeramente lo vamos a realizar mediante **Cadena de Markov**.

El “problema” es que para poder usar Cadena de Markov, la suma de las filas tiene que ser igual a uno, con lo que necesitamos transformar la matriz (como hicimos en el ejemplo inicial).

Lo que vamos a realizar para poder usar Cadena de Markov es sumar cada una de las filas y a este número restar cada elemento de la matriz (inicial), y este sería el numerador del elemento de la nueva matriz y lo vamos a dividir entre el denominador de la nueva matriz, que lo obtendremos sumando los elementos de las filas de la nueva matriz.

Vamos a explicar cómo se calcularía uno, y los demás se realizan de forma análoga. El elemento $a_{1,2}$ de la nueva matriz saldría de la siguiente manera: Sumamos la primera fila, 6+4+9+…+7 que es igual a . A este 55 le restamos el elemento a_{12} de la matriz de partida y obtenemos el 49, que será el numerador del elemento a_{12} de la nueva matriz. Hacemos lo mismo con el elemento $a_{1,3}$, $a_{1,4}$, …y así hasta el $a_{1,15}$ y después sumamos la fila de la nueva matriz, el número que obtenemos es el nuevo denominador de esta fila, 495.

La nueva matriz pues, nos quedaría tal como se muestra a continuación:

	1	2	3	4	5	6	7	8	9	10	11
1	0.00000000	0.0989899	0.1030303	0.09292929	0.09494949	0.1050505	0.1010101	0.09090909	0.1090909	0.1070707	0.09696970
2	0.10101010	0.0000000	0.1070707	0.09696970	0.09494949	0.0989899	0.1030303	0.09090909	0.1050505	0.1090909	0.09292929
3	0.09696970	0.1030303	0.0000000	0.09292929	0.09090909	0.0989899	0.1050505	0.09494949	0.1070707	0.1090909	0.10101010
4	0.09696970	0.1050505	0.1030303	0.00000000	0.09494949	0.1010101	0.0989899	0.09090909	0.1070707	0.1090909	0.09292929
5	0.09696970	0.1030303	0.1050505	0.09292929	0.00000000	0.0989899	0.1010101	0.09090909	0.1070707	0.1090909	0.09494949
6	0.10303030	0.0969697	0.1050505	0.09494949	0.09292929	0.0000000	0.1010101	0.09090909	0.1090909	0.1070707	0.09898990
7	0.09696970	0.1010101	0.1090909	0.09292929	0.09090909	0.0989899	0.0000000	0.09494949	0.1070707	0.1050505	0.10303030
8	0.09292929	0.1010101	0.1090909	0.09494949	0.09090909	0.0969697	0.1030303	0.00000000	0.1070707	0.1050505	0.09898990
9	0.10303030	0.1010101	0.1050505	0.09090909	0.09494949	0.1070707	0.0989899	0.09292929	0.0000000	0.1090909	0.09696970
10	0.10101010	0.1070707	0.1090909	0.09696970	0.09494949	0.1030303	0.0989899	0.09090909	0.1050505	0.0000000	0.09292929
11	0.09696970	0.1010101	0.1090909	0.09494949	0.09090909	0.0989899	0.1050505	0.09292929	0.1070707	0.1030303	0.00000000

Analizamos un poco la matriz. Cada peña a la que pertenece el caballo (cliente) tiene su opinión sobre la clasificación de los Caballos del Vino, vamos por filas. Por ejemplo, para el caballo número 1 (recordamos que nos estamos refiriendo a la peña que es la que vota), el mejor sería el 9, que es TERRY (probabilidad más alta), y el peor sería el número 8 SOLTERON-TRIANA. En la segunda fila, el caballo núero 2 ya opina diferente, ya que para él, el mejor sería el caballo número 10, UNIVERSO, y coloca en última posición al caballo número 8, SOLTERON-TRIANA. Y si seguimos analizando los opiniones por filas, ocurrirá lo mismo, no se ponen todos de acuerdo en cual tiene que ser la clasificafión idonea de los caballos. Luego ya tenemos nuestro problema.

Para resolverlo, al igual que en el ejemplo inicial, vamos a usar el software R Studio, con el paquete “markovchain”.

```
install.packages("markovchain")
library(markovchain)
#TFG
estados<-c("1","2","3","4","5","6","7","8","9","10","11")
matrizTransicion=matrix(data=c(0,49/495,51/495,46/495,47/495,52/495,50/495,45/495,54/495,53/495,48/495,
                               50/495,0,53/495,48/495,47/495,49/495,51/495,45/495,52/495,54/495,46/495,
                               48/495,51/495,0,46/495,45/495,49/495,52/495,47/495,53/495,54/495,50/495,
                               48/495,52/495,51/495,0,47/495,50/495,49/495,45/495,53/495,54/495,46/495,
                               48/495,51/495,52/495,46/495,0,49/495,50/495,45/495,53/495,54/495,47/495,
                               51/495,48/495,52/495,47/495,46/495,0,50/495,45/495,54/495,53/495,49/495,
                               48/495,50/495,54/495,46/495,45/495,49/495,0,47/495,53/495,52/495,51/495,
                               46/495,50/495,54/495,47/495,45/495,48/495,51/495,0,53/495,52/495,49/495,
                               51/495,50/495,52/495,45/495,47/495,53/495,49/495,46/495,0,54/495,48/495,
                               50/495,53/495,54/495,48/495,47/495,51/495,49/495,45/495,52/495,0,46/495,
                               48/495,50/495,54/495,47/495,45/495,49/495,52/495,46/495,53/495,51/495,0),
                        byrow=TRUE,nrow=11,dimnames=list(estados,estados))
CadenaMarkov=new("markovchain",states=estados,byrow=TRUE,transitionMatrix=matrizTransicion,name="Caballos del Vino")
show(CadenaMarkov)
plot(CadenaMarkov)
#APARTADO C
steadyStates(CadenaMarkov) #estado equilibrio
```

Mediante el estado de equilibrio vemos la clasificación óptima:

```
> steadyStates(CadenaMarkov) #estado equilibrio
              1          2          3          4          5          6          7          8          9        10        11
[1,] 0.08982491 0.09242409 0.09623475 0.08603768 0.08521374 0.09163925 0.09223203 0.08437417 0.09670098 0.0969036 0.0884148
```

Si la ordenamos estos valores obtenemos que la clasificación óptima de los Caballos del Vino es:

<u>CLASIFICACIÓN ÓPTIMA CABALLOS DEL VINO (CADENA MARKOV)</u>

POSICIÓN FINAL	NOMBRE CABALLO
1	UNIVERSO
2	TERRY
3	CAPRICHOSO
4	CAMPEON
5	SANGRINO
6	MEL-AZULES
7	AMBICIOSO
8	ZAMBRA
9	ESPARTACO
10	JUPITER
11	SOLTERON-TRIANA

Vemos que el Caballo ganador sería UNIVERSO, en segundo lugar estaría TERRY, y completando el podio está CAPRICHOSO. Por la cola los tres caballos con menor puntuación o menor coeficiente son ESPARTACO, JUPITER Y SOLTERON-TRIANA, por ese orden.

A continuación, vamos a pasar a plantear y resolver el problema mediante el otro método visto, **<u>Ordenamiento Óptimo (Linear Ordering Problem).</u>**

Como en el anterior, volvemos a partir de la matriz inicial:

$$\begin{pmatrix} 0 & 6 & 4 & 9 & 8 & 3 & 5 & 10 & 1 & 2 & 7 \\ 5 & 0 & 2 & 7 & 8 & 6 & 4 & 10 & 3 & 1 & 9 \\ 7 & 4 & 0 & 9 & 10 & 6 & 3 & 8 & 2 & 1 & 5 \\ 7 & 3 & 4 & 0 & 8 & 5 & 6 & 10 & 2 & 1 & 9 \\ 7 & 4 & 3 & 9 & 0 & 6 & 5 & 10 & 2 & 1 & 8 \\ 4 & 7 & 3 & 8 & 9 & 0 & 5 & 10 & 1 & 2 & 6 \\ 7 & 5 & 1 & 9 & 10 & 6 & 0 & 8 & 2 & 3 & 4 \\ 9 & 5 & 1 & 8 & 10 & 7 & 4 & 0 & 2 & 3 & 6 \\ 4 & 5 & 3 & 10 & 8 & 2 & 6 & 9 & 0 & 1 & 7 \\ 5 & 2 & 1 & 7 & 8 & 4 & 6 & 10 & 3 & 0 & 9 \\ 7 & 5 & 1 & 8 & 10 & 6 & 3 & 9 & 2 & 4 & 0 \end{pmatrix}$$

Recordamos que este método es un sistema de ordenación que se basa en sumar los elementos por encima de la diagonal principal de una matriz, y se le llama Ordenamiento Óptimo, al que da la suma mayor.

Necesitamos todas las matrices resultantes de todas las posibles permutaciones de filas y columnas, basándonos en la puntuación que cada peña a la pertenece un Caballo del Vino hace a los demás caballos.

Para saber el ordenamiento óptimo, tenemos que considerar todas las permutaciones posibles, en este caso, como hay 11 caballos, tendremos 11! matrices (ordenamientos o permutaciones posibles).

Explicamos un poco la primera inicial, el resto se interpretaran igual.

	AMBICIOSO	CAMPEON	CAPRICHOSO	ESPARTACO	JUPITER	MEL-AZULES	SANGRINO	SOLTERON-TRIANA	TERRY	UNIVERSO	ZAMBRA
AMBICIOSO	0	6	4	9	8	3	5	10	1	2	7
CAMPEON	5	0	2	7	8	6	4	10	3	1	9
CAPRICHOSO	7	4	0	9	10	6	3	8	2	1	5
ESPARTACO	7	3	4	0	8	5	6	10	2	1	9
JUPITER	7	4	3	9	0	6	5	10	1	2	6
MEL-AZULES	7	5	1	9	10	6	0	8	2	3	4
SANGRINO	9	5	1	8	10	7	4	0	2	3	6
SOLTERON-TRIANA	4	5	3	10	8	2	6	9	0	1	7
TERRY	5	2	1	7	8	4	6	10	3	0	9
UNIVERSO	7	5	1	8	10	6	3	9	2	4	0
ZAMBRA	0	6	4	9	8	3	5	10	1	2	7

Cada peña a la que pertenece el caballo es como si fuese un cliente, y tiene su opinión sobre cuál es el ordenamiento óptimo. En la primera fila, el caballo número 1, AMBICIOSO “dice” que el mejor es el 10, UNIVERSO (por eso lo coloca en primera posición dándole el 1), y el peor es el número 8, SOLTERON-TRIANA (por eso le da la posición un 10). En la segunda fila, la peña a la que pertenece el caballo número 2, CAMPEON, ya opina diferente, y si seguimos analizando lo que opinan todas las peñas a las pertenecen los caballos por filas, no habría acuerdo, luego volvemos a tener un problema.

Para resolverlo vamos a utilizar el RStudio, la sintaxis o código que utilizamos es el siguiente:

```r
install.packages("gtools")
library(gtools)

vector <-seq(11)
ordenes<-permutations(11,11,vector)


lop=0
final<-factorial(11)
mejor<-c(1:final)
mejorf=0
for( i in 1:final)
{
  obj=0
  for (j in 1:11)
  {
    if (j<11)
    {
      for (k in (j+1):11)
      {
        obj=obj+A[ordenes[i,j],ordenes[i,k]]
      }
    }
  }
  mejor[i]=obj
  if (obj>mejorf)
  {
    mejorf=obj
    lop=i
  }
}
mejorf
lop
definitivo<-c(1:11)

for(i in 1:11)
{
definitivo[i]=ordenes[lop,i]
}
definitivo
```

R nos devuelve el valor de la mayor suma de los elementos por encima de la diagonal principal de todas las posibles permutaciones de la matriz que hemos introducido anteriormente, la fila donde se encuentra esa mayor suma, y por último nos imprime esa fila, que es el ordenamiento óptimo de los Caballos del Vino que buscábamos.

```
> mejorf
[1] 374
> lop
[1] 10234461
> definitivo<-c(1:11)
> for(i in 1:11)
+ {
+ definitivo[i]=ordenes[lop,i]
+ }
> definitivo
 [1]  3 10  2  9  7  6  1 11  5  4  8
```

Observamos que la suma mayor de las puntuaciones es 374, la cual se encuentra en la fila 10234461, y que corresponde con la siguiente clasificación:

CLASIFICACIÓN ÓPTIMA CABALLOS DEL VINO (LOP)

POSICIÓN FINAL	NOMBRE CABALLO
1	CAPRICHOSO
2	UNIVERSO
3	CAMPEON
4	TERRY
5	SANGRINO
6	MEL-AZULES
7	AMBICIOSO
8	ZAMBRA
9	JUPITER
10	ESPARTACO
11	SOLTERON-TRIANA

Observamos que según Ordenamiento Óptimo (Linear Ordering Problem) el caballo ganador sería CAPRICHOSO, seguido en segunda posición de UNIVERSO, y completando él podía estaría CAMPEON. En las últimas posiciones estarían JUPITER, ESPARTACO y SOLTERON-TRIANA, por ese orden.

Ahora vamos a comparar la clasificación que obtuvimos con Cadena de Markov, con lo que nos ha proporcionado el Ordenamiento Óptimo (Linear Ordering Problem), y así vemos más claras las diferencias que encontramos entre una clasificación y la otra.

CADENA DE MARKOV		ORDENAMIENTO ÓPTIMO (LOP)	
POSICIÓN FINAL	**NOMBRE DEL CABALLO**	**POSICIÓN FINAL**	**NOMBRE DEL CABALLO**
1	**UNIVERSO**	1	**CAPRICHOSO**
2	**TERRY**	2	**UNIVERSO**
3	**CAPRICHOSO**	3	**CAMPEON**
4	**CAMPEON**	4	**TERRY**
5	SANGRINO	5	SANGRINO
6	MEL-AZULES	6	MEL-AZULES
7	AMBICIOSO	7	AMBICIOSO
8	ZAMBRA	8	ZAMBRA
9	**ESPARTACO**	9	**JUPITER**
10	**JUPITER**	10	**ESPARTACO**
11	SOLTERON-TRIANA	11	SOLTERON-TRIANA

Vemos que hay diferencias con respecto a la parte alta de la clasificación, y también por abajo.

Si usamos Cadena de Markov para ordenar los Caballos del Vino, el ganador sería UNIVERSO (queda segundo en LOP), mientras que en Ordenamiento Óptimo el ganador sería CAPRICHOSO (que queda en tercera posición en Markov). TERRY estaría segundo en Markov, mientras que usando LOP estaría cuarto, y CAMPEON está cuarto en Cadena de Markov, mientras que finaliza tercero en LOP.

Por abajo lo único que nos cambia de una clasificación a otra, es que ESPARTACO y JUPITER cambian sus posiciones, mientras que ESPARTACO ocupa la posición antepenúltima en Markov, en LOP está en la penúltima. Y JUPIETER se encuentra en la penúltima en Markov, y en la antepenúltima en LOP.

6. SIMILITUDES Y DIFERENCIAS

De cara a elegir un método u otro, decir que no existe uno mejor, sino dependiendo de las circunstancias, ambos pueden ser válidos y de mucha utilidad.

De forma general, podemos afirmar que ambos métodos pueden tener empates. Cadena de Markov es una media ponderada en la que tienen más peso los ganadores.

Ambos métodos son más sofisticados que por ejemplo usar las medias, pues obtendríamos muchos más empates. Para las medias sumariamos las columnas de la matriz.

Una diferencia general es que usando Ordenamiento Óptimo lo que buscamos para ser el mejor es el número que sea más pequeño, sin embargo en Cadena de Markov es justo lo contrario, el mejor es el mayor.

En los ejemplos estudiados, hemos visto los dos casos posibles, en el del fútbol, tanto por Cadena de Markov como por Ordenamiento Óptimo obteníamos el mismo resultado, la clasificación no nos variaba, con lo que resulta indiferente elegir entre un método u otro.

Sin embargo en el ejemplo estudiado de los Caballos del Vino de Caravaca de la Cruz, obteníamos resultados distintos, la clasificación final nos variaba en la parte alta sobretodo y en la baja también, entonces, ¿qué método es mejor aquí?

Deberíamos fijarnos en que ventajas y desventajas nos proporciona cada método y utilizar el que más nos interese. Si no se pueden producir empates nos interesa más elegir Ordenamiento Óptimo (Linear Ordering Problem), aunque como desventaja podemos decir que aquí importan más los ganadores, en nuestro ejemplo de los Caballos del Vino si gana UNIVERSO, lo que haya dicho este tiene más peso.

Ordenamiento Óptimo se diferencia de Cadena de Markov en que es un acuerdo entre todos, donde todos tienen el mismo peso, pero pueden producirse empates.

En nuestro ejemplo de los Caballos del Vino, la mejor solución es elegir Cadena de Markov, porque no se pueden producir empates.

7. CONCLUSIONES

En el mundo actual existen muchos casos reales en los que se precisa de un sistema de ordenación de votaciones o puntuaciones y necesitamos tomar la decisión de cuál es el óptimo.

Hemos estudiado dos métodos para ordenar sistemas de votaciones (entre los múltiples que hay), Cadena de Markov (Markov Chain) y Ordenamiento Óptimo (Linear Ordering Problem), y los hemos comparado mediante ejemplos.

Lo que hemos aprendido con este trabajo es:

- El método de unificación de votaciones Ordenamiento Óptimo (Linear Ordering Problem), que es y cómo aplicarlo. Como implementarlo en el software Matlab y Rstudio.
- Revisión de la bibliografía del método Cadena de Markov, el cual ya fue estudiado en tercero del grado de estadística empresarial.
- Aplicación de ambos métodos a un caso real, la fiesta de los Caballos del Vino de Caravaca de la Cruz.
- Mejora en el uso del programa RStudio.

Vamos a ver ahora como se relaciona el trabajo con el grado:

Asignatura	**Tema**
Simulación de Procesos y Sistemas	Cadenas de Markov y Código RStudio
Modelos de Optimización	Optimización lineal y Código Matlab
Introducción a la Programación	Código RStudio
Gestión de Cartera e Inversiones	Código RStudio

8. REFERENCIAS BIBLIOGRAFICAS

- *ORDESHOOK, PETER C.: GAME THEORY AND POLITICAL THEORY. AND INTRODUCTION. Cambridge University Press, 1999.*

  http://asojodcr.blogspot.com.es/2009/03/la-paradoja-de-condorcet.html

- *The Linear Ordering Problem: Methods in Combinatorial Optimization (Applied Mathematical Sciences 175) del autor S.S. Antman J.E. Marsden L. Sirovich. (página 7)*
- *Campos, V., Glover, F. Laguna, M. and Martí, R.: An experimental evaluation of a scatter search for the linear ordering problem, Journal of Global Optimizationz 21 (2001), 397–414.*
- *García, C.G., Pérez-Brito, D., Campos, V. and Martí, R.: Variable neighborhood search for the linear ordering problem, Computers and Operations Research 33 (2006),3549–3565*
- *THE LINEAR ORDERING PROBLEM: Exact and Heuristic Methods in Combinatorial Optimization (Applied Mathematical Sciences 175).*
- *http://www.optsicom.es/lolib/*

- *GONZALEZ CARMONA, ANDRES, SAENZ BARRIO OSCAR: CADENAS DE MARKOV APLICADAS AL DIAGNÓSTICO Y TRATAMIENTO DE CONDUCTAS Y PROCESOS DE APRENDIZAJE, 32-45*
- *UTILIZAR LAS CADENAS DE MARKOV PARA TOMAR MEJORES DECISIONES. Tirado, Maribel. 22/01/2104 (Artículo original de Michael Walker en Data Science Association).*

  http://www.bigdatahispano.org/noticias/cadenas-de-markov-para-tomar-mejores-decisiones/

- *http://www.sitiosdeespana.es/Evento/caballos-del-vino#.WRsWqWjyhPY*
- *Charon, I. and Hudry, O.: A survey on the linear ordering problem for weighted or unweighted tournaments, 4OR 5 (2007), 5–60.*