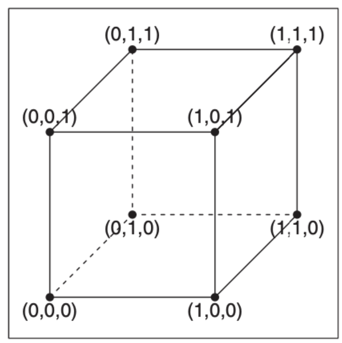
EJERCICIOS DEL PRIMER BLOQUE ***CURSO 2021-2022***
Cadenas de Markov

1. Estudia las propiedades de la cadena de Markov (X_n) cuya matriz de transiciones viene dada por
$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$$
Representa dicha cadena de Markov con un grafo.

2. Considera el grafo del cubo,



cuyos vértices consideramos los estados de una cadena de Markov cuya matriz de transición es

$$\boldsymbol{P} = \begin{array}{c} \\ 000 \\ 100 \\ 010 \\ 110 \\ 001 \\ 101 \\ 011 \\ 111 \end{array} \begin{array}{c} \begin{array}{cccccccc} 000 & 100 & 010 & 110 & 001 & 101 & 011 & 111 \end{array} \\ \begin{pmatrix} 0 & 1/3 & 1/3 & 0 & 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 0 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 1/3 & 0 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 1/3 & 0 & 1/3 & 1/3 & 0 \end{pmatrix} \end{array}.$$

- Interpreta el significado de la cadena de Markov así definida.
- Estudia las propiedades de dicha cadena
- Realiza, usando SAGE, una simulación de la cadena y explica los resultados a la luz de los que has obtenido en los apartados anteriores.

3. El índice meteorológico canadiense de incendios forestales se utiliza ampliamente como un medio para estimar el riesgo de incendios forestales. El Ministerio de Recursos Naturales de Ontario utiliza el índice para clasificar el riesgo diario de incendios forestales como nulo, bajo, moderado, alto o extremo. Martell (1999) recopiló datos diarios de riesgo de incendio durante 26 años en 15 estaciones meteorológicas en Ontario para construir un modelo de cadena de Markov de cinco estados para los cambios diarios en el índice. La matriz de transición de una ubicación para la subsección de principios de verano es

$$\boldsymbol{P} = \begin{array}{l} \\ \text{Nil} \\ \text{Low} \\ \text{Moderate} \\ \text{High} \\ \text{Extreme} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Nil} & \text{Low} & \text{Moderate} & \text{High} & \text{Extreme} \end{array} \\ \begin{pmatrix} 0.575 & 0.118 & 0.172 & 0.109 & 0.026 \\ 0.453 & 0.243 & 0.148 & 0.123 & 0.033 \\ 0.104 & 0.343 & 0.367 & 0.167 & 0.019 \\ 0.015 & 0.066 & 0.318 & 0.505 & 0.096 \\ 0.000 & 0.060 & 0.149 & 0.567 & 0.224 \end{pmatrix} \end{array}.$$

Para los administradores forestales es de interés la distribución de probabilidad a largo plazo del índice diario. Independientemente del riesgo en un día en particular, ¿cuál es la probabilidad de riesgo a largo plazo para un día típico a principios del verano?

4. Considere un paseo aleatorio en $\{0, ..., k\}$, que se mueve hacia la izquierda y hacia la derecha con las respectivas probabilidades q y p. Si el camino está en 0, pasa a 1 en el siguiente paso. Si el camino está en k, se mueve a $k-1$ en el siguiente paso. A esto se le llama camino aleatorio con límites reflectantes. Suponga que $k = 3$, $q = 1/4$, $p = 3/4$ y que la distribución inicial es uniforme.

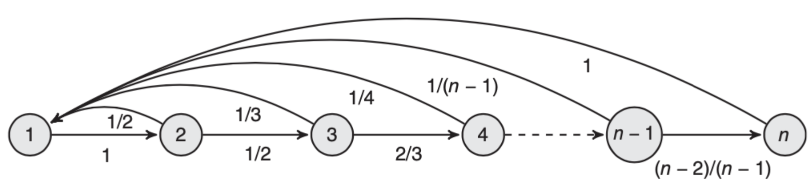
   - Exhibe la matriz de transición.
   - Encuentra $P(X_7 = 1|X_0 = 3, X_2 = 2, X_4 = 2)$.
   - Encuentra $P(X_3 = 1, X_5 = 3)$.

5. Un dado en forma de tetraedro tiene cuatro caras etiquetadas $1, 2, 3$ y 4. En tiradas independientes repetidas del dado $R_0, R_1, ...$, sea $X_n = \max\{R_0, ..., R_n\}$ el valor máximo después de $n+1$ lanzamientos del dado, para $n \geq 0$.

   - Dé un argumento intuitivo de por qué $X_0, X_1, ...$, es una cadena de Markov, y exhibe la matriz de transición.
   - Encuentra $P(X_3 \geq 3)$.

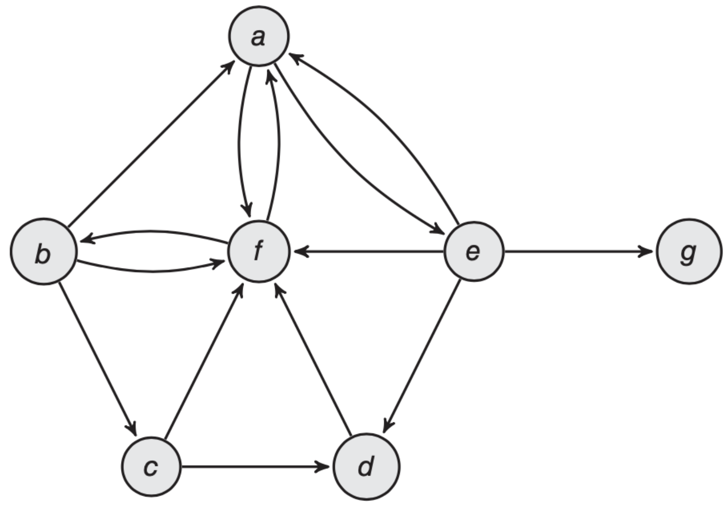
6. Sea $X_0, X_1, ...$ una cadena de Markov con matriz de transición P. Sea $Y_n = X_{3n}$, para $n = 0, 1, 2, ...$ Demuestra que $Y_0, Y_1, ...$ es una cadena de Markov y exhibe su matriz de transición.

7. Considera la cadena de Markov cuyo grafo de transición asociado es



   Estudia las propiedades de la cadena de Markov asociada.

8. (PageRank) El algoritmo de búsqueda de PageRank de Google se basa en el modelo de navegación aleatoria, que es una caminata aleatoria en el gráfico de la web. Para este gráfico, cada vértice representa una página de Internet. Una arista dirigida conecta i con j si hay un enlace de hipertexto de la página i a la página j. Cuando el usuario está en la página i, se mueve a una nueva página eligiendo entre los enlaces disponibles en i de manera uniforme y aleatoria. La siguiente figura muestra una red simplificada con siete páginas.



La red está descrita por la matriz:

$$
\boldsymbol{N} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \\ f \\ g \end{array}
\begin{array}{c} \begin{array}{ccccccc} a & b & c & d & e & f & g \end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 \\
1/3 & 0 & 1/3 & 0 & 0 & 1/3 & 0 \\
0 & 0 & 0 & 1/2 & 0 & 1/2 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1/4 & 0 & 0 & 1/4 & 0 & 1/4 & 1/4 \\
1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} \end{array}.
$$

Tenga en cuenta que N no es una matriz estocástica, ya que la página g no tiene enlaces de salida, de modo que la fila asociada a g consta está formada completamente por ceros (y no suma 1). Para asegurarse de que el camino aleatorio llegua a todas las páginas de la red, el algoritmo debe tener en cuenta (i) las páginas que no tienen enlaces de salida, llamadas nodos colgantes, y (ii) los grupos de páginas que pueden hacer que el camino se atasque en un subgrafo. En la red del ejemplo, g es un nodo colgante. Suponga que la red consta de k páginas. En el algoritmo de PageRank, la solución para los nodos colgantes es asumir que cuando el usuario aleatorio llega a una página de este tipo, salta a una nueva página en la red de manera uniforme y aleatoria. Se obtiene una nueva matriz Q donde cada fila de N correspondiente a un nodo colgante se cambia por una en la que todas las entradas son $1/k$. La nueva matriz Q es una matriz estocástica. Para el problema de atascarse potencialmente en pequeños subgrafos de la red, la solución propuesta en el artículo original por Brin y Page (1998) fue fijar un factor de amortiguamiento $0 < p < 1$ para modificar la matriz Q. En su modelo, desde una página dada, el internauta aleatorio, con probabilidad $1 - p$, decide no seguir ningún enlace en la página y, en cambio, navegar a una nueva página aleatoria en la red. Por otro lado, con probabilidad p, sigue los enlaces de la página como de costumbre. Esto define la matriz de transición de PageRank

$$P = pQ + (1 - p)A,$$

donde A es una matriz $k \times k$ cuyas entradas son todas $1/k$. El factor de amortiguación utilizado por Google se estableció originalmente en $p = 0,85$. Con el factor de amortiguación, la matriz de PageRank P es estocástica y el camino aleatorio resultante es aperiódico e irreducible. El PageRank de una página en la red es la probabilidad estacionaria de esa página. Hacer los cálculos correspondientes con el grafo que se ha mostrado en este ejercicio.