



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
POSGRADO EN CIENCIAS MATEMÁTICAS

**"Solución numérica a procesos de Markov con espacio de estados finito: Una revisión teórica con aplicaciones"**

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
MAESTRO EN CIENCIAS

PRESENTA:  
RAFAEL MARÍN SALGUERO

DIRECTOR DE LA TESIS  
DRA. ÚRSULA ITURRARÁN-VIVEROS,  
[FACULTAD DE CIENCIAS \(UNAM\)](#)

MÉXICO, D. F. 25 NOVIEMBRE DE 2013



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

# Solución numérica a procesos de Markov con espacio de estado finito: Una revisión teórica con aplicaciones

Rafael Marín Salguero

E-Mail: [ramasa79@gmail.com](mailto:ramasa79@gmail.com)

# Contenido

<b>Contenido</b>	<b>1</b>
<b>1 Teoría básica de probabilidad</b>	<b>4</b>
1.1 Elementos básicos . . . . .	4
1.2 Variables aleatorias . . . . .	11
1.2.1 Distribuciones conjuntas de probabilidad . . . . .	14
1.3 Distribución Poisson y Exponencial . . . . .	17
<b>2 Teoría básica de matrices</b>	<b>22</b>
2.1 Función de una matriz y la matriz exponencial . . . . .	23
2.2 Método de la diagonalización . . . . .	31
2.3 Método de la uniformización . . . . .	39
2.4 Método de Padé . . . . .	50
2.5 Método de Padé mejorado . . . . .	59
<b>3 Teoría básica de procesos de Markov</b>	<b>67</b>
3.1 Cadenas de Markov finitas . . . . .	67
3.1.1 Cadenas irreducibles . . . . .	80
3.1.2 Cadenas Reducibles . . . . .	87
3.2 Procesos de Markov Finitos . . . . .	93
3.2.1 Procesos Irreducibles . . . . .	97
3.2.2 Procesos Reducibles . . . . .	101
<b>4 Algunos modelos y aplicaciones</b>	<b>107</b>
4.1 Redes estocásticas . . . . .	107
4.2 Teoría de Colas . . . . .	118
4.3 El problema de inclusión . . . . .	136
4.3.1 El logaritmo de una matriz . . . . .	136
4.3.2 Condiciones de inclusión . . . . .	139
4.3.3 Ajustes a la matriz logaritmo . . . . .	147
4.3.4 Un caso de estudio. Avance académico estudiantil . . . . .	154
4.4 Conclusiones . . . . .	166
<b>Bibliografía</b>	<b>168</b>

# Agradecimientos

Este trabajo hubiera sido imposible de realizar y concluir si no fuera por el apoyo incondicional de toda mi familia, es difícil expresar con palabras todas las emociones que hemos vivido durante estos últimos años. Solamente me queda agradecer toda la paciencia y cariño que han tenido para mi. Le dedico a mi tía Mary y a mi sobrino Andrés este trabajo, que descansan en paz.

A mis amigos que siempre estuvieron apoyándome en todo momento, en especial a Bety por impulsarme a llevar a cabo este proyecto. Gracias con todo cariño.

A mis sinodales M. en C. María Lourdes Velasco Arregui, Dr. Jesús López Estrada, Dr. Francisco Javier Domínguez Mota y al Dr. Pedro González Casanova por todos los comentarios y sugerencias que surgieron en la revisión de este trabajo, y ayudaron en la mejora del manuscrito.

En especial quiero agradecer de todo corazón a la Dra. Úrsula Iturrarán-Viveros por todo su apoyo incondicional que recibí en la última etapa de la maestría y durante todo el desarrollo de este trabajo, gracias por ser una parte importante en este proyecto.

# Resumen

Este trabajo esta dedicado a realizar un análisis numérico sobre los diferentes métodos que existen para calcular las probabilidades de transición de un proceso de Markov con espacio de estado discreto. Para ello en el primer capítulo se mostraron los resultados básicos de la teoría de probabilidad necesarios, como la probabilidad condicional, fundamental en el estudio de las cadenas y procesos de Markov. En el segundo capítulo se hizo una revisión general de la teoría de matrices, la definición de la matriz exponencial y la demostración de algunas de sus propiedades más importantes, así como los diferentes métodos que existen para cálculo de funciones de matrices, en particular de la matriz exponencial (método de diagonalización, uniformización, método de Padé). En el tercer capítulo se enumeraron elementos básicos generales de la teoría tanto de las cadenas de Markov como de los procesos de Markov, mostrando resultados importantes para ambos procesos estocásticos y, a su vez, la similitud que existen entre ellos. Por último, en el cuarto capítulo se vieron algunas aplicaciones importantes dentro de los procesos de Markov como son las redes estocásticas, la teoría de colas y el problema de inclusión (cálculo de la matriz logaritmo). En la exposición del trabajo se mostraron pequeños códigos realizados en Matlab, que pueden ser utilizados también en el programa de uso libre Octave, para la facilitación del análisis, la comparación de distintos métodos y el mejor entendimiento de la teoría. Sin embargo, estos códigos están pensados solamente para uso académico, y se pueden modificar para obtener resultados más concretos o ampliarlos.

# Prefacio

El objetivo de este trabajo es mostrar las herramientas matemáticas básicas sobre las cadenas y procesos de Markov con espacio de estados finito y una revisión básica sobre los distintos métodos para calcular las probabilidades de transición para cada  $t \in \mathbb{R}$ , dadas por:

$$\text{a) } \frac{d\mathbf{P}(t)}{dt} = \mathbf{A}\mathbf{P}(t) \quad \text{con } \mathbf{P}(0) = \mathbf{I}$$

$$\text{b) } \mathbf{P}(t) = e^{\mathbf{A}t} \quad \text{con } t \in \mathbb{R}^+.$$

Para ello, se utilizarán inicialmente dos referencias básicas, el libro de Higham [8] es una excelente referencia sobre los distintos métodos que hay actualmente para calcular funciones de matrices y el libro de Stewart [13] que trata sobre los métodos numéricos para la solución de cadenas y procesos de Markov finitos. Con estas dos referencias básicas se desarrollarán códigos elaborados en los paquetes de cómputo especializados Matlab<sup>1</sup> y/o Octave<sup>2</sup> para obtener información relevante de los procesos de Markov. También se compara cuál de los distintos métodos para encontrar las probabilidades de transición  $\mathbf{P}(t) = e^{\mathbf{A}t}$  es más eficiente, numéricamente más estable y aplicarlo en la resolución de ciertos problemas básicos que surgen de los distintos modelos que hay sobre los procesos de Markov. Se busca que este trabajo sea accesible para personas que no tengan una formación de matemático, pero que les permita entender los conceptos básicos sobre los procesos de Markov y usar las herramientas computacionales desarrollados en Matlab y/o Octave. Se decidió trabajar con Matlab (u Octave) porque el ambiente de programación facilita mucho los cálculos numéricos que se requieren. Su lenguaje de programación es sencillo de entender y aplicar, ya que están especializados para trabajar con matrices, y tienen la ventaja de que es fácil entender lo que hace el código. También su forma vectorial de programación lo hace muy eficiente frente a otros programas computacionales como Mathematica o Maple. Dentro de este trabajo se estará haciendo referencia a Mat/Oct, para indicar que en cualquiera de los dos programas los códigos se ejecutan de manera correcta. Todos los cálculos computacionales se realizaron en una PC con procesador Intel Core i5 de 64 bits, con una velocidad del CPU de 2.67 Ghz y una memoria RAM de 6GB.

---

<sup>1</sup>Matlab © Es propiedad de The Mathworks Inc. Todos los derechos reservados.

<sup>2</sup>Octave ©. Desarrollado por John W. Eaton. Su uso se autoriza bajo la licencia GNU General Public License. Su página de distribución es <http://www.gnu.org/software/octave/>

# Capítulo 1

## Teoría básica de probabilidad

### 1.1 Elementos básicos

Mostraremos los conceptos básicos que se necesitarán en el transcurso de este trabajo sobre la teoría de probabilidad y de los procesos estocásticos. En la práctica se tienen fenómenos aleatorios donde uno tiene la necesidad de poder cuantificar en la medida de lo posible la incertidumbre de cierto evento. Se trabajará con los elementos de las cadenas de Markov con espacio de estados finito y tiempo finito, y con los procesos de Markov a tiempo continuo pero con espacio de estados finitos. La teoría de la probabilidad fue desarrollada con rigor matemático, en un principio, por el matemático ruso Andréi Nikoláyevich Kolmogorov<sup>1</sup> en el año de 1933. Los procesos de Markov son probabilidades condicionales inicialmente desarrolladas por el matemático ruso Andrei Andreyevich Markov<sup>2</sup> en el año de 1907. En este primer capítulo se hará referencia a los resultados básicos de la teoría de la probabilidad para entender los procesos de Markov y sus aplicaciones. A los fenómenos donde no se conoce de antemano el resultado cuando se observa una realización se les denomina *aleatorios*. El término *fenómeno aleatorio*, se entenderá como cualquier experimento real que pueda dar lugar a varios resultados sin que sea posible anunciar con certeza cuál de estos resultados va a ser observado. Veamos que elementos están involucrados en un fenómeno aleatorio.

**Definición 1.1** *Al espacio constituido por la colección de todos los posibles resultados diferentes de un fenómeno aleatorio se le conoce **espacio muestral** y será denotado por el*

---

<sup>1</sup> Andrei Nikolaevich Kolmogorov (1903-1987). Matemático ruso. En el año de 1933 publicó "*Foundations of the Theory of Probability*" (*Fundamentos de la teoría de probabilidad*) donde da la primera versión de su construcción matemática rigurosa de la probabilidad. Publicó más de 200 artículos sobre teoría de funciones, lógica matemática, teoría de las probabilidades, teoremas límite, teoría de los algoritmos, educación e historia de las matemáticas entre otros.

<sup>2</sup> Andrei Andreyevich Markov (1856-1922). Matemático ruso. En un principio sus trabajos fueron sobre teoría de números, límites de integrales, análisis y convergencia de series. A partir del año 1900 empezó a estudiar sucesiones de variables aleatorias dependientes, esperando establecer leyes de probabilidad en su forma más general. El demostró el teorema del límite central bajo condiciones muy generales. A Markov se le recuerda particularmente por sus estudios sobre *cadenas*, que son sucesiones de variables aleatorias en donde el futuro es determinado por el presente, pero es independiente en la forma en que se llegó a este. Disfrutaba mucho de la poesía, y es por eso que la primera aplicación de las cadenas fuera estudiar la distribución de consonantes y vocales del poema Eugene Onegin de A. S. Pushkin. Fue hasta el año de 1926 cuando el matemático ruso S. N. Bernstein utilizó la frase *cadenas de Markov*.

símbolo  $\Omega$  (que puede ser numerable o no).

Con este espacio muestral  $\Omega$  sea  $\mathcal{A} \subset 2^\Omega$  una clase de subconjuntos del **conjunto potencia** de  $\Omega$ , que de manera práctica se puede pensar que  $\mathcal{A}$  son los eventos observables del fenómeno aleatorio.

**Definición 1.2** Una clase de conjuntos  $\mathcal{A} \subset 2^\Omega$  es llamada la  $\sigma$ -álgebra, si cumple con las siguientes propiedades:

a)  $\Omega \in \mathcal{A}$

b)  $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$

c)  $A_1, A_2, A_3, \dots \in \mathcal{A} \Rightarrow \bigcup_{k=1}^{\infty} A_k \in \mathcal{A}$

**Definición 1.3** Un suceso o evento  $A$  es un subconjunto de  $\Omega$  tal que  $A \in \mathcal{A}$ , donde  $\mathcal{A}$  es una  $\sigma$ -álgebra de  $\Omega$ .

Al querer conocer cuál es la probabilidad de que ocurra un cierto evento, entendiendo que la probabilidad "mide" la incertidumbre de un evento de un fenómeno aleatorio dado. La construcción básica de la probabilidad de un evento está dada por

$$P[\text{evento } A] = \frac{\# \text{ casos favorables } A}{\# \text{ casos totales}}. \quad (1.1)$$

a esta expresión se le conoce como la **probabilidad clásica** de un evento. En un experimento aleatorio con pocos resultados es sencillo enumerar todos los posibles resultados del fenómeno y poder contar los casos favorables, pero en la práctica esto resulta imposible debido a que el espacio muestral puede ser muy complicado o imposible de encontrar (por no ser finito), y para poder encontrar una probabilidad en cuestión uno tiene que hacer uso de la probabilidad frecuencial ó simulación probabilista. La idea es poder repetir un experimento aleatorio y verificar la ocurrencia o no de nuestro evento a medir, y realizar el cociente de "casos favorables entre casos totales". Se puede definir la probabilidad frecuencial de la siguiente manera

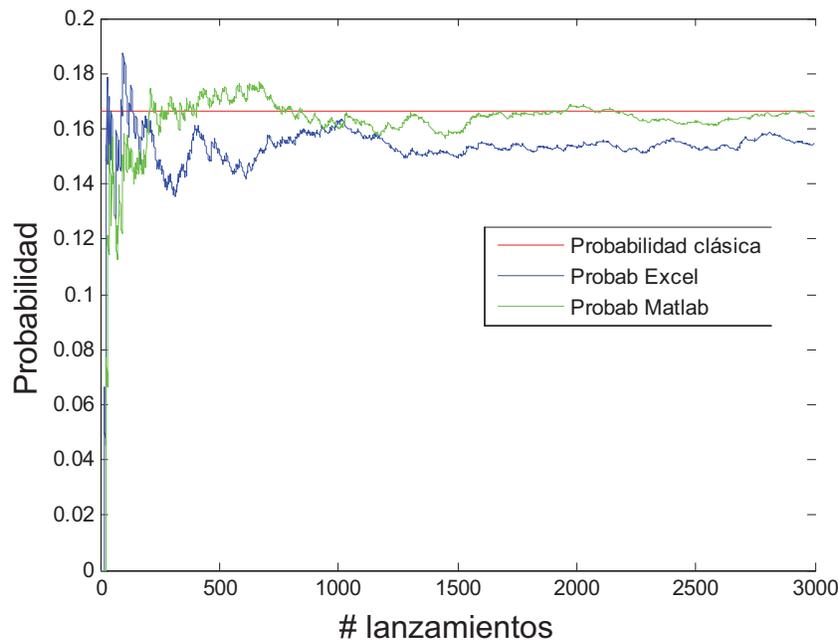
$$\tilde{P}_N[\text{evento } A] = \frac{\sum_{k=1}^N \mathbf{1}_{\{X_k=A\}}}{N}. \quad (1.2)$$

donde  $\mathbf{1}_{\{X_k=A\}}$  denota a la **función indicadora** y vale 1 si el evento  $A$  se observó en la  $k$ -ésima realización del fenómeno y 0 si no se observó, y  $N$  es el número de realizaciones del fenómeno. Asumiendo que si los eventos del fenómeno aleatorio se realizan bajo las

mismas condiciones entonces la probabilidad frecuencial converge a la probabilidad clásica del evento  $A$ .

$$\lim_{N \rightarrow \infty} \tilde{P}_N[A] = P[A]. \quad (1.3)$$

A continuación se muestra un ejemplo en el cual el experimento aleatorio consiste en lanzar dos dados honestos y observar si la suma de los dados es menor o igual que 4. Se realizó una simulación donde se muestra la diferencia entre estas dos probabilidades, un bloque de  $N = 2998$  lanzamientos se hicieron tanto en Mat/Oct como en la hoja de cálculo comercial Excel, para mostrar la diferencia entre los dos sistemas.



Con esto se puede observar que el generador de números pseudo aleatorios<sup>3</sup> es mucho mejor en Mat/Oct que el que se maneja en la hoja de cálculo, por lo que se recomienda evitar en la medida de lo posible la realización de simulaciones en estos programas no optimizados para la simulación numérica. También es importante subrayar que el método de simulación probabilista es eficiente para el cálculo de probabilidades complicadas.

**Proposición 1.4** Una función de probabilidad  $P : \mathcal{A} \rightarrow [0, 1]$  donde  $\mathcal{A}$  es una  $\sigma$ -álgebra de  $\Omega$ , cumple las siguientes propiedades:

a)  $P[\emptyset] = 0$

<sup>3</sup>El generador de números pseudo aleatorios que trae los programas de cómputo actuales es el conocido como "Twister-Mersenne", desarrollado por los matemáticos Makoto Matsumoto y Takuji Nishimura [16]. La propiedad fundamental del algoritmo es que el generador se repetirá después  $2^{19937} - 1$  valores, en términos prácticos son  $2.15771 \times 10^{6001}$  valores. Es decir, si hicieramos 1,000,000 de números pseudo-aleatorios cada segundo en un año calcularíamos obtendríamos  $3.1536 \times 10^{13}$  números, y estos se repetirán aproximadamente en  $6.84206 \times 10^{5987}$ , lo que es realmente sorprendente. En la práctica no tendremos porque preocuparnos por la generación de números pseudo-aleatorios.

- b)  $0 \leq P[A] \leq 1, \forall A \in \mathcal{A}$
- c) Para cualquier  $A, B \in \mathcal{A} \Rightarrow P[A \cup B] = P[A] + P[B] - P[A \cap B]$
- d) Si  $A \in \mathcal{A}$  tenemos que  $P[A^c] = 1 - P[A]$
- e) Si  $A_1, A_2, A_3, \dots \in \mathcal{A}$  tal que  $\forall i \neq j, A_i \cap A_j = \emptyset$  se cumple que

$$P \left[ \bigcup_{i=1}^{\infty} A_i \right] = \sum_{i=1}^{\infty} P[A_i].$$

Estas propiedades básicas de la función de probabilidad nos indican verbalmente que la ocurrencia del evento vacío o imposible tiene una probabilidad cero de ocurrir, pero es importante subrayar que el regreso no siempre es cierto, es decir, pueden existir eventos  $A \in \mathcal{A}$  cuya probabilidad sea cero pero no imposibilita su ocurrencia. La segunda propiedad nos indica que la probabilidad de cualquier evento tiene una probabilidad entre 0 y 1. Entre más cercana a cero nos dirá que es un evento "raro", que es poco probable observar su ocurrencia. En el otro caso una probabilidad cercana a 1 nos dirá que es un evento que ocurre con mucha frecuencia al repetir varias veces el fenómeno aleatorio. La tercera propiedad nos indica que si queremos calcular la probabilidad de la ocurrencia del evento  $A$  ó del evento  $B$ , sumamos las probabilidades de la ocurrencia de cada evento por separado y quitaremos la probabilidad de la ocurrencia de ambos simultáneamente. En general esta propiedad se generaliza de la siguiente manera:

$$P \left[ \bigcup_{k=1}^N A_k \right] = \sum_{k=1}^N P[A_k] - \sum P[\cap \text{eventos 2 a 2}] + \sum P[\cap \text{eventos 3 a 3}] + (-1)^{N+1} P \left[ \bigcap_{k=1}^N A_k \right]. \quad (1.4)$$

La cuarta propiedad se refiere a que la probabilidad de la no ocurrencia del evento  $A$  es lo mismo que la resta de la unidad menos la probabilidad del evento en cuestión. Hay algunos problemas prácticos donde es más fácil calcular la probabilidad del complemento que del evento original en cuestión. Por último, si se tiene una colección finita o numerable de eventos disjuntos, entonces la probabilidad de la suma de los eventos es igual a la suma de las probabilidades de los eventos individuales. Para ver elementos un poco más teóricos sobre la teoría general de probabilidad se recomiendan; revisar el material expuesto en García [3], Feller [20], Klenke [15] o Billingsley [26].

Otro de los componentes muy importantes dentro del estudio de la probabilidad es el probabilidad condicional, se refiere a que actualmente se sabe que ocurrió un cierto evento y se está interesado en conocer la ocurrencia de subeventos implicados por la ocurrencia de un evento. En la literatura matemática actual a estas probabilidades se le denota como  $P[A | B]$  y se lee como la probabilidad del evento  $A$  dado que ocurrió el evento  $B$ . En el trabajo de Kolmogorov denotaba a estas probabilidades como  $P_B[A]$ , pero es una notación

que no ha sido totalmente adoptada. A continuación se definirán ciertas propiedades de estas probabilidades condicionales y algunos elementos importantes como es la probabilidad total, regla de Bayes<sup>4</sup>, entre otras.

**Definición 1.5** Sean  $A$  y  $B$  dos eventos tal que  $P[B] > 0$ . Se define la **probabilidad condicional de  $A$  dado  $B$**  como

$$P[A | B] = \frac{P[A \cap B]}{P[B]}. \quad (1.5)$$

De esta definición se puede reescribir la probabilidad de la intersección de dos eventos

$$P[A \cap B] = P[A | B]P[B] = P[B | A]P[A]. \quad (1.6)$$

y dependiendo de que lado de la igualdad se utilice se pide que  $P[B] > 0$  o  $P[A] > 0$ .

**Teorema 1.6 (Probabilidad Total)** Sea  $\Omega$  un espacio muestral y sea  $\{B_n\}_{n \in \mathbb{N}}$  una **partición**<sup>5</sup> de  $\Omega$ . Entonces, para cualquier evento  $A$  se obtiene

$$P[A] = \sum_{i \in \mathbb{N}} P[A \cap B_i]. \quad (1.7)$$

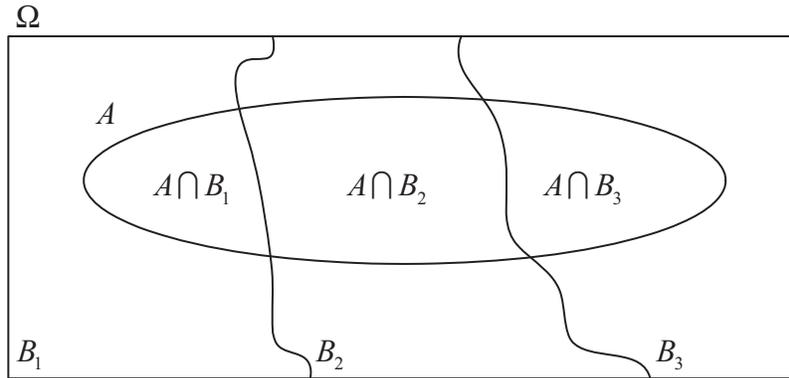
si además se cumple que  $P[B_i] > 0 \forall i \in \mathbb{N}$  se satisface

$$P[A] = \sum_{i \in \mathbb{N}} P[A | B_i]P[B_i]. \quad (1.8)$$

<sup>4</sup>Thomas Bayes (1702-1761). Matemático británico. Miembro de la Royal Society desde 1742, Bayes fue uno de los primeros en utilizar la probabilidad inductivamente y establecer una base matemática para la inferencia probabilística. En 1763, dos años después de su muerte, se publica *Essay Towards Solving a Problem in the Doctrine of Chances*, en el que trataba el problema de las causas a través de los efectos observados, y donde se enuncia el teorema que lleva su nombre. El trabajo fue entregado a la Royal Society por Richard Price y es la base de la técnica bayesiana.

<sup>5</sup>Debemos recordar que una partición de  $\Omega$  es una colección de conjuntos  $\{B_n\}_{n \in \mathbb{N}}$  que satisface las siguientes relaciones:

(1)  $B_j \neq \emptyset$  para  $j \in \mathbb{N}$ , (2)  $B_i \cap B_j = \emptyset$  para  $i \neq j$ , (3)  $\bigcup_{j \in \mathbb{N}} B_j = \Omega$

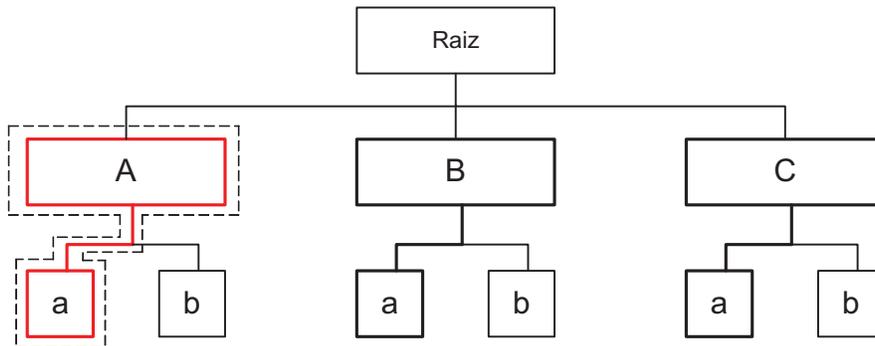


Interpretacion de la probabilidad total para  $n = 3$

$$P[A] = P[A \cap B_1] + P[A \cap B_2] + P[A \cap B_3]$$

**Teorema 1.7 (Teorema de Bayes)** *Sea  $\Omega$  un espacio muestral y sea  $\{B_n\}_{n \in \mathbb{N}}$  una partición de  $\Omega$  tal que  $P[B_i] > 0$ . Entonces, para cualquier evento  $A$  que cumpla con  $P[A] > 0$  se obtiene*

$$P[B_j | A] = \frac{P[A | B_j]P[B_j]}{\sum_{i \in \mathbb{N}} P[A | B_i]P[B_i]} \tag{1.9}$$



Interpretacion de la probabilidad

$$\begin{aligned} P[A | a] &= \frac{P[A \cap a]}{P[a]} = \frac{P[A \cap a]}{P[a \cap A] + P[a \cap B] + P[a \cap C]} \\ &= \frac{P[a | A]P[A]}{P[a | A]P[A] + P[a | B]P[B] + P[a | C]P[C]} \end{aligned} \tag{1.9}$$

y en este caso se utilizan las propiedades de (1.7) y (1.6)

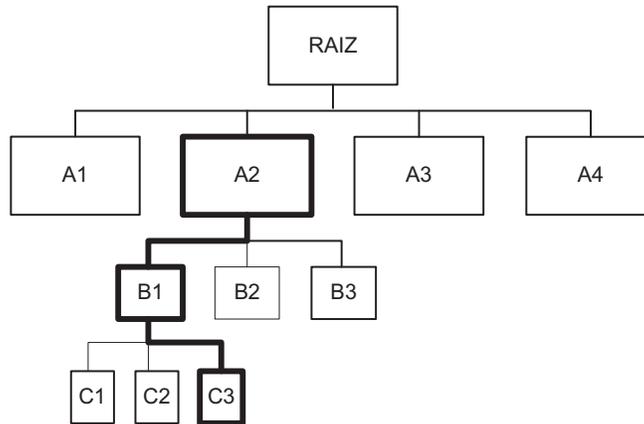
**Definición 1.8** Una familia  $\{A_k\}_{k \in I}$  de eventos indexado por un conjunto arbitrario  $I$  es llamada **independiente** si para cualquier colección finita de eventos distintos  $A_{i_1}, \dots, A_{i_k}$  de esta familia se cumple que

$$P \left[ \bigcap_{j=1}^k A_{i_j} \right] = \prod_{j=1}^k P[A_{i_j}]. \tag{1.10}$$

En probabilidad dos ideas importantes con la intersección de dos eventos, los cuáles son el de eventos ajenos y eventos independientes. Dos eventos  $A$  y  $B$  son **ajenos** (o mutuamente excluyentes) si se cumple que  $P[A \cap B] = 0$  y son **independientes** si  $P[A \cap B] = P[A]P[B]$ . Cuando de antemano se sabe que los eventos no son independientes, entonces se puede aplicar lo que se conoce como la **regla de la cadena** ó **regla del producto**<sup>6</sup>

$$P \left[ \bigcap_{j=1}^n A_j \right] = P \left[ A_n \mid \bigcap_{j=1}^{n-1} A_j \right] \cdots P[A_2 \mid A_1] P[A_1]. \tag{1.11}$$

y tendría la siguiente interpretación geométrica (que es como si calculáramos la probabilidad de una rama en un "árbol probabilístico")



Interpretación de la probabilidad

$$P [C_3 \cap B_1 \cap A_2] = P [C_3 \mid B_1 \cap A_2] P [B_1 \mid A_2] P [A_2]$$

<sup>6</sup>También se le conoce como la **regla del producto**, esto debido a que se puede confundir con la regla de la cadena del cálculo diferencial para derivar funciones compuestas

$$\frac{d}{dx} [f(g(x))] = f'(g(x))g'(x)$$

Esta última relación se obtiene a partir de usar dos veces la relación (1.6)

$$\begin{aligned} P[A \cap B \cap C] &= P[C | A \cap B]P[A \cap B] \\ &= P[C | A \cap B]P[B | A]P[A]. \end{aligned}$$

## 1.2 Variables aleatorias

**Definición 1.9** Un *espacio de probabilidad* es una terna  $(\Omega, \mathcal{A}, P)$  donde  $\Omega$  es el espacio muestral,  $\mathcal{A}$  es una  $\sigma$ -álgebra de  $\Omega$  y  $P$  es una función de probabilidad definida sobre  $\mathcal{A}$ .

**Definición 1.10** Sea  $(\Omega, \mathcal{A}, P)$  un espacio de probabilidad. Una *variable aleatoria* es una función  $X : \Omega \rightarrow \mathbb{R}$  tal que  $\forall x \in \mathbb{R}$  el conjunto  $\{X \leq x\}$  sea un evento, esto es

$$\boxed{\{X \leq x\} \in \mathcal{A}.} \quad (1.12)$$

Una función  $X : \Omega \rightarrow \mathbb{B}$ , donde  $\Omega$  y  $\mathbb{B}$  son conjuntos finitos o a lo más numerables es llamada *variable aleatoria discreta* si  $\forall x \in \mathbb{B}$

$$\{X = x\} \in \mathcal{A}.$$

**Proposición 1.11** Si  $(\Omega, \mathcal{A}, P)$  es un espacio de probabilidad y  $X$  es una variable aleatoria, entonces  $\forall x, y \in \mathbb{R}$  las siguientes condiciones definen a una variable aleatoria

$$\boxed{\begin{aligned} a) \{X < x\} &\in \mathcal{A} \\ b) \{X \geq x\} &\in \mathcal{A} \\ c) \{X > x\} &\in \mathcal{A} \\ d) \{x \leq X \leq y\} &\in \mathcal{A}. \end{aligned}} \quad (1.13)$$

En la literatura hay dos funciones importantes asociadas a la variable aleatoria  $X$ .

**Definición 1.12** Sea  $(\Omega, \mathcal{A}, P)$  un espacio de probabilidad y  $X$  una variable aleatoria definida sobre ese espacio. La función

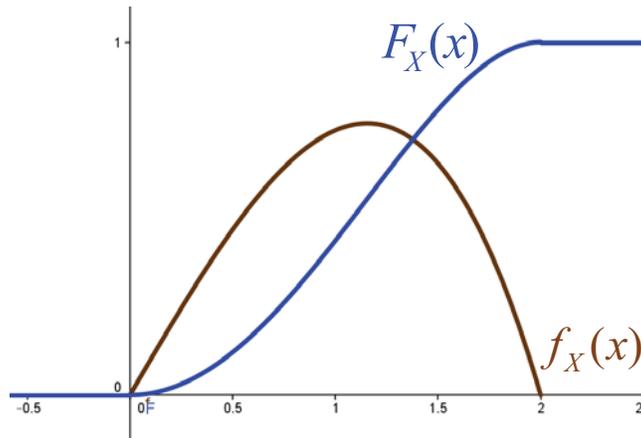
$$\boxed{F_X(x) = P[X \leq x] \quad \forall x \in \mathbb{R}.} \quad (1.14)$$

se le conoce como la *función de distribución* de  $X$ . También se define a la función  $f_X(x) = P[X = x]$  como la *función de densidad* de  $X$  y ambas se relacionan mediante  $F_X(x) = \sum_{u \leq x} f_X(u)$

En García [3] vienen demostradas las siguientes propiedades básicas de la función de densidad y de distribución:

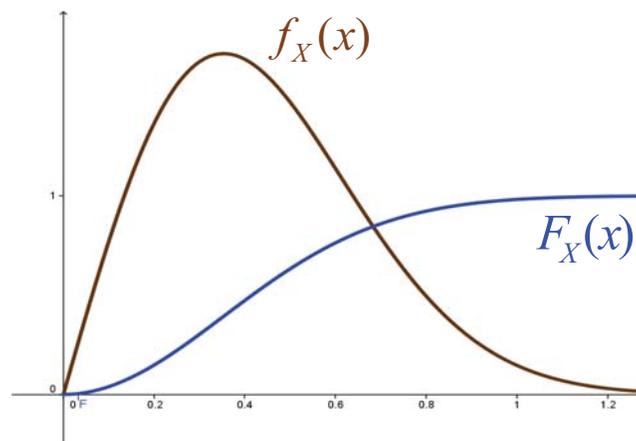
$$\begin{aligned}
 &\text{a) } F_X(x) \text{ es una función monótona no decreciente y continua por la derecha} \\
 &\text{b) } \lim_{x \rightarrow -\infty} F_X(x) = 0 \text{ y } \lim_{x \rightarrow \infty} F_X(x) = 1 \\
 &\text{c) } f_X(x) = F_X(x) - F_X(x^-) \\
 &\text{d) } f_X(x) \geq 0, \text{ si es discreta la v.a. } X \text{ entonces } \sum_{x=-\infty}^{\infty} f_X(x) = 1 \\
 &\quad \text{si es continua la v.a. } X \text{ entonces } \int_{\mathbb{R}} f_X(x) dx = 1 \\
 &\text{e) } F_X(x) = \int_{-\infty}^x f_X(u) du \text{ si la función de densidad } f_X \text{ es absolutamente continua.} \\
 &\text{f) } P[a < X \leq b] = F_X(b) - F_X(a) \text{ para } a < b.
 \end{aligned}
 \tag{1.15}$$

Estas propiedades matemáticas permiten construir funciones de densidad y de distribución de manera matemática sin que sea necesario (inicialmente) definir un fenómeno aleatorio. También, en el caso continuo, si se tiene la expresión matemática de la función de distribución  $F_X(x)$  entonces la función de densidad está dada por  $f_X(x) = \frac{dF_X}{dx}$ . Algunos ejemplos de estas funciones de densidad se muestran a continuación.



$$f_X(x) = \begin{cases} x - \frac{x^3}{4} & x \in [0, 2] \\ 0 & \text{e.o.c} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{-1}{16}(x^4 - 8x^2) & x \in [0, 2] \\ 1 & \text{e.o.c} \end{cases}$$



$$f_X(x) = \begin{cases} 8xe^{-4x^2} & x \in [0, \infty) \\ 0 & \text{e.o.c} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-4x^2} & x \in [0, \infty) \end{cases} .$$

Estos son algunos ejemplos de funciones de densidad y de distribución de variables aleatorias  $X$  de algún fenómeno aleatorio en particular, por el momento "no es necesario" especificar algún fenómeno aleatorio. Es importante señalar que en particular hay dos elementos importantes asociados a las funciones de densidad (o de distribución) que son la **esperanza** (media ó promedio)  $E[X]$  y la **varianza**  $\text{Var}[X]$ . La media nos da información de cuáles son

los valores que en promedio toma una variable aleatoria y la varianza nos da información de la dispersión que existe de los datos alrededor de la media. Sus definiciones matemáticas son:

	$X$ discreta	$X$ continua
<b>Media</b>	$E[X] = \sum_x xP[X = x]$	$E[X] = \int_D x f_X(x) dx$
<b>Varianza</b>	$\text{Var}[X] = \sum_x (x - \mu)^2 P[X = x]$	$\text{Var}[X] = \int_D (x - \mu)^2 f_X(x) dx.$

(1.16)

donde  $\mu = E[X]$  es el promedio de la variable aleatoria. En la varianza se toma el cuadrado de la diferencia debido a que la función es continua y derivable en  $x = 0$ , por eso se evita tomar el valor absoluto de la diferencia. Otro elemento de la varianza es que las diferencias grandes las hace más grandes y las diferencias pequeñas las hace más pequeñas, por lo que en ocasiones conviene trabajar más con la **desviación estándar**, que es la raíz cuadrada de la varianza. Si  $\sigma^2 = \text{Var}[X]$  entonces  $\sigma = \sqrt{\text{Var}[X]}$  representa la desviación estándar de la variable aleatoria. Algunas propiedades básicas que se pueden derivar de la media y varianza se muestran a continuación:

<b>Media</b>	<b>Varianza</b>
a) $E[X + Y] = E[X] + E[Y]$	a) $\text{Var}[X + \beta] = \text{Var}[X]$ con $\beta \in \mathbb{R}$
b) $E[aX + b] = aE[X] + b$ con $a, b \in \mathbb{R}$	b) $\text{Var}[\beta X] = \beta^2 \text{Var}[X]$
c) $E[c] = c$ con $c \in \mathbb{R}$	c) $\text{Var}[X] = E[X^2] - E^2[X].$

(1.17)

### 1.2.1 Distribuciones conjuntas de probabilidad

Las propiedades anteriores serán importantes para poder analizar de manera básica cadenas de Markov continuas con espacio de estados finito. Todas las propiedades vistas anteriormente son para funciones de variables aleatorias con una sola variable. A continuación se mostrarán algunos resultados básicos de las variables aleatorias conjuntas. Para dos variables aleatorias  $X, Y$  sobre un mismo espacio  $\Omega$  de probabilidad se define la **función de densidad conjunta**  $f_{X,Y}(x, y)$  de la siguiente manera

$$f_{X,Y}(x, y) = P[X = x \cap Y = y] = P[X = x, Y = y]$$

(1.18)

Algunas propiedades elementales que tiene esta función de densidad conjunta es que

$f_{X,Y}(x, y) \geq 0$  y además

Caso discreto	Caso continuo
$\sum_{x \in D_x} \sum_{y \in D_y} f_{X,Y}(x, y) = 1$	$\int_{D_x} \int_{D_y} f_{X,Y}(x, y) dy dx = 1.$

A partir de la función de densidad conjunta  $f_{X,Y}(x, y)$  se construye la **función de distribución conjunta**  $F_{X,Y}(x, y) = P[X \leq x, Y \leq y]$  y cumple lo siguiente

Caso discreto	Caso continuo	
$F_{XY}(x, y) = \sum_{s \leq x} \sum_{t \leq y} f_{XY}(s, t)$	$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(s, t) ds dt.$	(1.19)

Estas definiciones son cuando tenemos dos variables aleatorias y su función de densidad general, y se desea obtener información conjunta, pero en general se puede definir la función de distribución conjunta para  $n$  variables aleatorias. Dado que se tiene la función de distribución conjunta, se puede construir la **función de distribución marginal** como

$F_Y(y) = \lim_{x \rightarrow \infty} F_{XY}(x, y) \text{ para } y \in \mathbb{R}$	(1.20)
$F_X(x) = \lim_{y \rightarrow \infty} F_{XY}(x, y) \text{ para } x \in \mathbb{R}.$	

**Proposición 1.13** *Si  $F_{XY}$  y  $f_{XY}$  representan la función de distribución conjunta y función de densidad conjunta de dos variables aleatorias  $X, Y$ , entonces se cumplen las siguientes propiedades:*

- a)  $F_X(x) = \sum_y F_{XY}(x, y)$  si  $F_{XY}$  es discreta
- b)  $f_X(x) = \sum_y f_{XY}(x, y)$  si  $f_{XY}$  es discreta
- c)  $F_X(x) = \int_{-\infty}^{+\infty} F_{XY}(x, y) dy$  si  $F_{XY}$  es continua
- d)  $f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy$  si  $f_{XY}$  es continua.

Dadas estas propiedades de las distribuciones conjuntas y marginales se definen las distribuciones condicionales. Supongamos que  $X, Y$  son distribuciones conjuntas de probabilidad. Asumiendo que por el momento las distribuciones conjuntas son discretas, y se observó

que la variable  $Y = y$  ¿qué información se puede obtener de la variable  $X$ ? Para responder se define la **función de distribución condicionada** dada por:

$$F_{X|Y}(x; y) = P[X \leq x | Y = y]. \quad (1.21)$$

Esto nos dice que se tendrá una variable aleatoria  $X$  dado que ya conocemos el valor observado de  $Y = y$ . Junto con la función de distribución condicionada se define la función de densidad condicionada dada por

$$f_{X|Y}(x; y) = P[X = x | Y = y] \quad (1.22)$$

Para aplicar de manera más práctica estas definiciones se debe recordar la definición de probabilidad condicional (1.5) y se obtienen las siguientes definiciones un poco más fáciles de aplicar<sup>7</sup>

$$F_{X|Y}(x; y) = \frac{P[X \leq x, Y = y]}{P[Y = y]} = \frac{\sum_{t \leq x} f_{XY}(t, y)}{f_Y(y)}$$

$$f_{X|Y}(x; y) = \frac{P[X = x, Y = y]}{P[Y = y]} = \frac{f_{XY}(x, y)}{f_Y(y)}.$$

aunque es necesario pedir la condición de que  $f_Y(y) \neq 0$  para evitar que haya ambigüedades en la distribución condicional. Análogamente se construye la distribución condicional  $F_{Y|X}$  y  $f_{Y|X}$ . Cuando la función de densidad de conjunta de las variables aleatorias  $X$  y  $Y$  es absolutamente continua, la función de distribución condicional dado  $Y = y$  se define como

$$F_{X|Y}(x; y) = \frac{\int_{-\infty}^x f_{XY}(s, y) ds}{f_Y(y)} \quad (1.23)$$

Se puede tener el caso de que las variables aleatorias  $X$  y  $Y$  sean independientes una de otra, pero para verificar que sean independientes debe ocurrir que si observamos un valor  $Y = y$ , entonces la función de distribución condicional  $F_{X|Y}(x; y)$  no debe depender del valor  $Y = y$  para cualquier  $y \in \mathbb{R}$ , (análogamente si se observó el valor de  $X = x$ ) por lo que se debe de cumplir

$$F_{X|Y}(x; y) = F_X(x)$$

$$F_{Y|X}(y; x) = F_Y(y).$$

<sup>7</sup>En la definición de probabilidad condicional de 2 eventos se tiene que  $P[A | B] = \frac{P[A \cap B]}{P[B]}$

Cuando se trabaja con distribuciones conjuntas se omite el símbolo de intersección  $\cap$  y se utiliza mejor la coma ", " para simplificar la notación.

es decir, que la distribución condicional es una función de distribución de una sola variable. Entonces, haciendo la analogía a la probabilidad de eventos, para que dos variables aleatorias sean **independientes** su distribución conjunta se puede escribir como el producto de distribuciones marginales

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

$$f_{XY}(x, y) = f_X(x)f_Y(y).$$

Estos elementos básicos de la teoría de la probabilidad serán suficiente para entender las técnicas básicas involucradas en los procesos de Markov, pero hay mucha literatura donde se pueden ver con mayor detalle estos elementos probabilísticos. Se recomiendan las siguientes referencias De Groot [7], Feller [20], Grimmett [11], Harris [6], García [3] y [4], Klenke [15] para profundizar sobre el estudio sobre la probabilidad. A continuación se analizarán brevemente dos distribuciones de probabilidad que son importantes y que aparecen de manera natural en los procesos de Markov: la distribución Poisson y la distribución exponencial.

### 1.3 Distribución Poisson y Exponencial

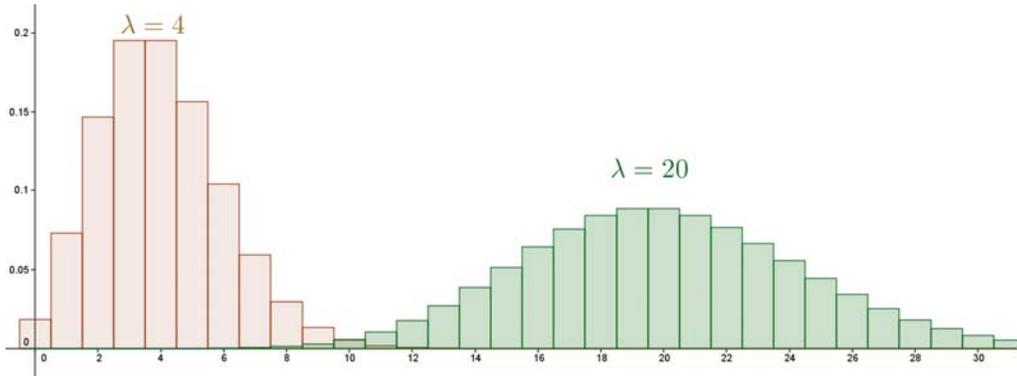
Hay dos distribuciones de probabilidad que son útiles en la teoría de los procesos estocásticos, la distribución Poisson<sup>8</sup> y la distribución Exponencial. Se dice que una variable aleatoria discreta  $X$  tiene una distribución Poisson con parámetro  $\lambda > 0$ , que se denota por  $X \sim \text{Poisson}(\lambda)$ , si su función de densidad está dada por:

$$P[X = k] = \frac{e^{-\lambda}\lambda^k}{k!} \text{ con } k = 0, 1, 2, \dots \quad (1.24)$$

Un significado práctico del parámetro  $\lambda$  es que representa el promedio de eventos que ocurren en cada unidad de tiempo, y este promedio permanece constante en intervalos disjuntos de tiempo. Esto quiere decir que la probabilidad de que un evento ocurra en una unidad fija de tiempo es el mismo para todas las unidades de tiempo. Dependiendo del contexto en el que se este trabajando, la tasa  $\lambda$  puede representar el número de bacterias en cultivos biológicos, el número de errores tipográficos en cada página de un libro, el número de personas que llegan a una central camionera, por mencionar algunos ejemplos, y es por esto que la distribución Poisson es muy útil en la teoría de la probabilidad. A continuación se muestran algunos ejemplos de los histogramas obtenidos con las probabilidades para distintos

<sup>8</sup>La distribución Poisson fue desarrollada por el matemático francés Siméon Denis Poisson (1781-1840). Fue en el año de 1837 cuando publicó su ensayo "Recherches Sur La Probabilité Des Jugements", en donde describe la ocurrencia de eventos con una probabilidad de éxito baja para que al repetirse en numerosas ocasiones el evento ocurre en algunas ocasiones. También Poisson realizó ensayos en astronomía, en la teoría del calor y en Mecánica, entre otros.

$\lambda$ , la cuál podemos pensar como una "tasa de intensidad".



Histogramas de las probabilidades de dos distribuciones Poisson con tasas  $\lambda = 4$  y  $\lambda = 20$

Esta representación gráfica nos dice que las probabilidades más altas se localizan cerca del parámetro  $\lambda$ , y las probabilidades empiezan a ser cada vez más pequeñas cuando uno se va alejando de la tasa  $\lambda$ . Esta propiedad viene demostrada en García [3], y se enuncia a continuación.

**Proposición 1.14** *Los términos  $B_k = \frac{e^{-\lambda}\lambda^k}{k!}$  de una distribución Poisson, crecen con  $k$  hasta alcanzar su máximo valor cuando  $\lambda - 1 \leq k \leq \lambda$  después decrecen con  $k$*

Otra propiedad importante que caracteriza a la distribución Poisson es que la esperanza y varianza de la distribución es igual a la tasa  $\lambda$ , esto se demuestra a continuación.

**Teorema 1.15** *Sea  $X \sim \text{Poisson}(\lambda)$  entonces la esperanza y varianza están dadas por*

$$E[X] = \lambda \quad \text{Var}[X] = \lambda. \tag{1.25}$$

**Demostración.** *Entonces, para calcular la esperanza usamos la definición (1.16)*

$$\begin{aligned} E[X] &= \sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} \\ &= \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!} = \lambda \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} = \lambda \sum_{m=0}^{\infty} \frac{e^{-\lambda}\lambda^m}{m!} \\ &= \lambda \end{aligned}$$

donde se utilizó el hecho de que  $n! = n(n-1)!$  y al hacer el cambio de variable la serie de nueva cuenta es una distribución Poisson. Para calcular la varianza se tiene que aplicar una técnica intermedia, que es calcular la esperanza  $E[X(X-1)]$  esto debido a que con ello

se cancelan dos términos factoriales y para ello se utiliza lo que se conoce como el **teorema del estadístico inconsciente**  $E[g(X)] = \sum g(x)P[X = x]$

$$\begin{aligned} E[X(X-1)] &= \sum_{k=0}^{\infty} k(k-1) \frac{e^{-\lambda} \lambda^k}{k!} = \sum_{k=2}^{\infty} k(k-1) \frac{e^{-\lambda} \lambda^k}{k!} \\ &= \sum_{k=2}^{\infty} \frac{e^{-\lambda} \lambda^k}{(k-2)!} = \lambda^2 \sum_{k=2}^{\infty} \frac{e^{-\lambda} \lambda^{k-2}}{(k-2)!} = \lambda^2 \sum_{m=0}^{\infty} \frac{e^{-\lambda} \lambda^m}{m!} \\ &= \lambda^2 \end{aligned}$$

Con esto se obtiene que  $E[X(X-1)] = E[X^2 - X] = E[X^2] - E[X] = \lambda^2$  entonces  $E[X^2] = \lambda^2 + \lambda$  por el resultado anterior, y utilizando las propiedades (1.17) queda como resultado

$$\text{Var}[X] = E[X^2] - E^2[X] = \lambda^2 + \lambda - \lambda^2 = \lambda$$

■

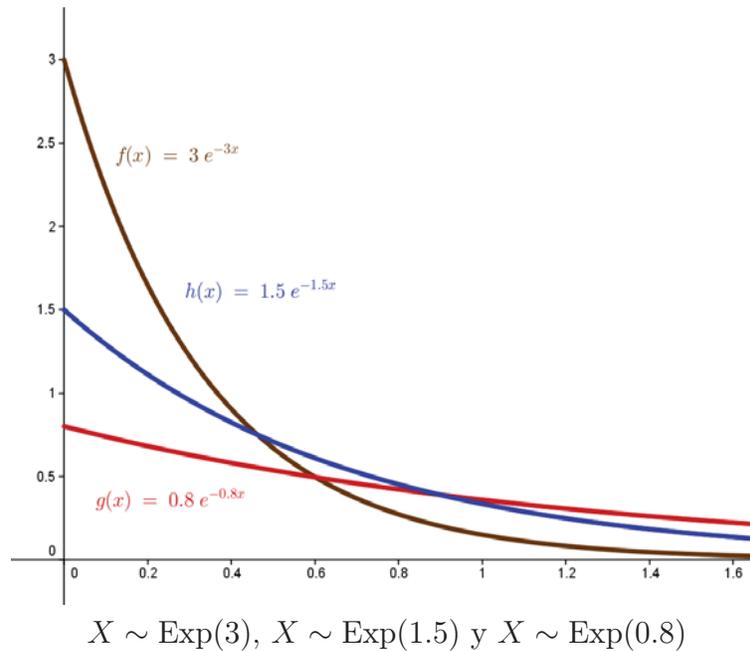
Estas propiedades de la distribución Poisson serán de utilidad más adelante al trabajar con lo que se conoce como la teoría de colas. También existen otras generalizaciones de la distribución de Poisson, que se pueden consultar en Feller [20], Medhi [21] ó Karlin [19] por citar algunos. Otra distribución muy importante en la teoría de la probabilidad es la **distribución exponencial**.

**Definición 1.16** Se dice que la variable aleatoria continua  $X$  tiene una **distribución exponencial** con parámetro  $\lambda > 0$ , y se denota por  $X \sim \text{Exp}(\lambda)$ , si tiene la siguiente función de densidad

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{e.o.c} \end{cases} \quad (1.26)$$

La distribución exponencial se puede obtener a partir de considerar el tiempo entre la realización de dos eventos consecutivos en el tiempo. Las gráficas de su función de densidad

se muestran a continuación.



Algunas propiedades básicas de la distribución exponencial se muestran a continuación.

**Proposición 1.17** Si  $X$  es una variable aleatoria con distribución exponencial entonces se cumplen las siguientes propiedades:

$$\begin{aligned}
 a) \quad E[X] &= \frac{1}{\lambda} \\
 b) \quad \text{Var}[X] &= \frac{1}{\lambda^2} \\
 c) \quad F_X(x) &= 1 - e^{-\lambda x} \text{ para } x \geq 0 \\
 d) \quad P[X \geq s + t \mid X \geq s] &= P[X \geq t]
 \end{aligned}
 \tag{1.27}$$

Las tres primeras propiedades son fáciles de demostrar haciendo uso de la integración por partes. Por la propiedad c) se tiene que  $P[X \geq x] = e^{-\lambda x}$ . A la propiedad d) se le conoce como la **pérdida de la memoria**. Para demostrarla se usará inicialmente la probabilidad condicional

$$\begin{aligned}
 P[X \leq s + t \mid X \geq s] &= \frac{P[s \leq X \leq s + t]}{P[X \geq s]} \\
 &= \frac{e^{-s\lambda}(1 - e^{-\lambda t})}{e^{-s\lambda}} \\
 &= 1 - e^{-\lambda t}
 \end{aligned}$$

Y calculando la probabilidad de el complemento de un evento se obtiene el resultado deseado

$$\begin{aligned}P[X \geq s + t \mid X \geq s] &= 1 - P[X \leq s + t \mid X \geq s] \\&= 1 - (1 - e^{-\lambda t}) \\&= e^{-\lambda t} = P[X \geq t]\end{aligned}$$

Es importante subrayar cuál es el significado práctico de la pérdida de la memoria. Supongamos que estamos esperando un transporte público, que ya hemos esperado más de media hora, la pregunta que nos podríamos hacer es ¿cuál es la probabilidad de que tengamos que esperar más de 45 minutos? Entonces, si definimos a la variable aleatoria  $X$  como el tiempo de espera, y asumiendo que se puede modelar como una distribución exponencial, esto lo podemos modelar como

$$P[X \geq 45 \mid X \geq 30] = P[X \geq 15]$$

es decir, se pierde el registro de que ya habíamos esperado media hora y esta probabilidad es equivalente a que tuviésemos que esperar más de 15 minutos de manera inicial. Con esto concluye un resumen general de los elementos básicos de probabilidad. En la siguiente sección se analizarán los elementos generales sobre la teoría de matrices, debido a que existe una relación matricial importante al querer estudiar características básicas sobre los procesos de Markov.

# Capítulo 2

## Teoría básica de matrices

Al trabajar con los procesos de Markov es necesario trabajar con las matrices, por lo que solamente a continuación se darán los elementos básicos así como la notación necesaria que se utilizará en este trabajo. Una matriz  $\mathbf{A}$  es un arreglo numérico de valores reales compuesto por  $nm$  elementos, distribuidos en  $n$  renglones y  $m$  columnas, que se denota por  $\mathbf{A} \in \mathbb{R}^{n \times m}$ . Las operaciones básicas se definen de manera distinta dependiendo si se trabaja con la suma o con la multiplicación. Para la suma de dos matrices  $\mathbf{A}$  y  $\mathbf{B}$  es necesario que ambas tengan las mismas dimensiones, y entonces la suma matricial se define elemento a elemento de la siguiente manera  $C_{i,j} = A_{i,j} \pm B_{i,j} \quad \forall i, j$  y de manera compacta  $\mathbf{C} = \mathbf{A} + \mathbf{B}$ . Con la multiplicación de dos matrices  $\mathbf{C} = \mathbf{AB}$  la única condición adicional que se pide para que este bien definida es que el número de columnas de la matriz  $\mathbf{A}$  sea igual al número de renglones de la matriz  $\mathbf{B}$ , es decir,  $\mathbf{A} \in \mathbb{R}^{n \times m}$  y  $\mathbf{B} \in \mathbb{R}^{m \times q}$ , y con esto la entrada  $C_{i,j}$  se define como

$$C_{i,j} = \sum_{k=1}^m A_{i,k} B_{k,j} \quad \forall i = 1, \dots, n \text{ y } j = 1, \dots, q. \quad (2.1)$$

Es importante subrayar que si no se cumple esta condición de igualdad entre las columnas de la primer matriz y los renglones de la segunda matriz entonces no esta definida la multiplicación. Algunas propiedades básicas de la suma y multiplicación, se muestran a continuación.

1) $\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$
2) $\lambda(\mathbf{A} + \mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B}$
3) $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$
4) $(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC}$
5) $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$
6) En general $\mathbf{AB} \neq \mathbf{BA}$ .

(2.2)

También es necesario trabajar con vectores, un vector renglón  $\mathbf{x} \in \mathbb{R}^{1 \times n}$  y un vector

columna  $\mathbf{y}' \in \mathbb{R}^{n \times 1}$  se definen de la siguiente manera

$$\mathbf{x} = [x_1 \quad x_2 \quad \cdots \quad x_n] \quad \mathbf{y}' = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}. \quad (2.3)$$

Cuando se trabaja con matrices, a la transpuesta de una matriz  $\mathbf{A}$ , que se denota por  $\mathbf{A}^T$ , intercambia los renglones por columnas y viceversa. Esto es importante debido a que en la literatura al vector columna generalmente se le denota por  $\mathbf{x}$  y al vector renglón como  $\mathbf{x}'$ , pero por cuestiones prácticas se decide utilizar la notación denotada en (2.3). Hay una literatura muy amplia sobre la teoría de matrices, las siguientes referencias son básicas Meyer [1], Davis [14], Cullen [17], Gantmacher[40] y Horn [45]. En ellas se enumeran y demuestran propiedades más generales sobre las matrices y vectores, y solamente se hará referencia a resultados importantes. Una parte importante dentro del álgebra lineal es el de poder definir funciones de matrices, una referencia fundamental en este tema está dado en Higham [8], y en nuestro caso solamente trabajaremos las propiedades de la matriz exponencial.

## 2.1 Función de una matriz y la matriz exponencial

A continuación se mostrarán los conceptos relacionados sobre la matriz exponencial. Una de las propiedades más importantes es la definición básica de funciones de matrices, que es una generalización de series infinitas de funciones de una sola variable.

**Definición 2.1** Una serie de potencias está definida por una serie convergente de la forma

$$\sum_{k=0}^{\infty} \beta_k x^k \text{ para } |x| < R \quad (2.4)$$

para una sucesión  $\{\beta_k\}_{k \in \mathbb{N}}$  y donde  $R$  es el radio de convergencia de la serie.

Con la serie de potencias se construyen series de potencias para funciones importantes dentro del análisis matemático como son la exponencial  $e^x$  y las funciones trigonométricas básicas  $\sin(x)$  y  $\cos(x)$

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots$$

Con estos resultados, existe un resultado similar para matrices cuadradas, que viene demostrado en Cullen [17], y da las condiciones generales que se deben cumplir para que se puede definir una función de una matriz mediante una serie de potencias.

**Teorema 2.2** Sea  $f(x)$  una función definida por una serie de potencias dada de la forma  $f(x) = \sum_{k=0}^{\infty} \beta_k x^k$  con  $|x| < R$  con  $R$  el radio de convergencia. Si  $\mathbf{A} \in \mathbb{R}^{n \times n}$  tiene su radio espectral<sup>1</sup>  $\rho(\mathbf{A}) < R$  entonces la serie de potencias matricial

$$f(\mathbf{A}) = \sum_{k=0}^{\infty} \beta_k \mathbf{A}^k. \quad (2.5)$$

converge.

Con este resultado importante se construye de manera general la exponencial de una matriz cuadrada, y esta definición está dada por

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k}{k!}. \quad (2.6)$$

y esta definición de la matriz exponencial trae consigo muchas propiedades básicas heredadas de la exponencial de una sola variable. En general este método de calcular la matriz exponencial mediante serie de potencias no es muy recomendable, en Moler-Van Loan [48], debido a que la matriz  $\mathbf{A}$  puede tener elementos grandes de signos contrarios y puede ocurrir errores numéricos grandes al hacer las restas. A continuación se mostrarán algunas propiedades que serán de mucha utilidad y están heredadas del caso escalar..

**Teorema 2.3** Sea  $\mathbf{F}(t) = e^{\mathbf{A}t}$  la matriz exponencial con  $t \in \mathbb{R}$  y sea  $s \in \mathbb{R}$  otro valor escalar. Entonces se cumple que

$$\mathbf{F}(t + s) = \mathbf{F}(t)\mathbf{F}(s). \quad (2.7)$$

**Demostración.** Lo que queremos demostrar es que se cumple en su forma matricial la conocida regla de los exponentes

$$e^{t+s} = e^t e^s$$

Para ello, tomaremos las expansiones en series de Taylor de la matriz exponencial

$$e^{\mathbf{A}t} e^{\mathbf{A}s} = \left( \mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \frac{\mathbf{A}^3 t^3}{3!} + \dots \right) \left( \mathbf{I} + \mathbf{A}s + \frac{\mathbf{A}^2 s^2}{2!} + \frac{\mathbf{A}^3 s^3}{3!} + \dots \right)$$

<sup>1</sup>Toda matriz cuadrada  $\mathbf{A}$  tiene  $\lambda_1, \lambda_2, \dots, \lambda_k$  valores propios distintos (contando multiplicidades) entonces el **radio espectral** de la matriz, denotado por  $\rho(\mathbf{A})$ , se define como el máximo de estos valores propios en valor absoluto.

$$\rho(\mathbf{A}) = \max \{ |\lambda_1|, |\lambda_2|, \dots, |\lambda_k| \}$$

compactando esta multiplicación de polinomios infinitos nos queda

$$e^{\mathbf{A}t}e^{\mathbf{A}s} = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{\mathbf{A}^{j+k} t^j s^k}{j!k!}$$

Haciendo un cambio de variable para la suma de los exponentes  $j + k = n$  y despejando a  $j$  obtenemos que  $j = n - k$  y por lo tanto

$$e^{\mathbf{A}t}e^{\mathbf{A}s} = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{\mathbf{A}^n t^{n-k} s^k}{(n-k)!k!} = \sum_{n=0}^{\infty} \mathbf{A}^n \sum_{k=0}^{\infty} \frac{t^{n-k} s^k}{(n-k)!k!}$$

La segunda suma tiene una forma incompleta. Recordando el desarrollo binomial obtenemos que

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k = \sum_{k=0}^n \frac{n!}{(n-k)!k!} x^{n-k} y^k$$

entonces, si a nuestra suma la multiplicamos y dividimos por  $n!$  obtendremos

$$\sum_{n=0}^{\infty} \frac{\mathbf{A}^n}{n!} \sum_{k=0}^{\infty} \frac{n! t^{n-k} s^k}{(n-k)!k!} = \sum_{n=0}^{\infty} \frac{\mathbf{A}^n}{n!} (t + s)^n = e^{\mathbf{A}(t+s)}$$

y por lo tanto

$$e^{\mathbf{A}t}e^{\mathbf{A}s} = e^{\mathbf{A}(t+s)}$$

■

El resultado anterior es muy importante, debido a que así podremos calcular las probabilidades de transición para tiempos más grandes, solamente tendremos que especializar el cálculo de la matriz exponencial en un solo tiempo fijo  $t^*$  y después solamente ir multiplicando esa matriz exponencial varias veces para obtener las probabilidades en tiempos mayores. Por ejemplo, si encontramos la matriz  $e^{0.1\mathbf{A}}$  usando algún método eficiente, entonces podremos calcular fácilmente las matrices  $e^{0.2\mathbf{A}}$  y  $e^{0.3\mathbf{A}}$  usando

$$\begin{aligned} e^{0.2\mathbf{A}} &= e^{0.1\mathbf{A}}e^{0.1\mathbf{A}} \\ e^{0.3\mathbf{A}} &= e^{0.1\mathbf{A}}e^{0.1\mathbf{A}}e^{0.1\mathbf{A}} \end{aligned}$$

y lo podremos lograr con relativa facilidad haciendo la multiplicación de matrices de manera reiterada. Otra derivación del resultado anterior es que se puede verificar rápidamente cuál es la inversa de la matriz exponencial. Si tomamos a  $s = -t$  obtenemos

$$e^{\mathbf{A}(0)} = e^{\mathbf{A}(t-t)} = e^{\mathbf{A}t}e^{\mathbf{A}(-t)} = e^{\mathbf{A}(-t)}e^{\mathbf{A}t} = e^{\mathbf{0}} = \mathbf{I}$$

pero esta es la condición para la construcción de la matriz inversa de  $e^{\mathbf{A}t}$ , por lo que obtenemos el siguiente resultado práctico, equivalente de la existencia del inverso multiplicativo de un número real  $\frac{1}{e^x} = e^{-x}$

$$\boxed{\boxed{(e^{\mathbf{A}t})^{-1} = e^{-\mathbf{A}t}}}$$

Recordando y utilizando el desarrollo del binomio de Newton

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k \text{ para } a, b \in \mathbb{R}. \quad (2.8)$$

podemos construir una expresión para la exponencial de dos matrices cuadradas cualesquiera  $e^{\mathbf{A}+\mathbf{B}}$ . Uno de los problemas básicos que se observa es que la multiplicación de matrices con  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$  es que en general no es conmutativa

$$\mathbf{AB} \neq \mathbf{BA}.$$

Podemos tomar dos ejemplos básicos para visualizar este hecho:

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 1 & 10 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} -2 & 4 \\ 0 & 2 \end{bmatrix}$$

$$\mathbf{AB} = \begin{bmatrix} -6 & 16 \\ -2 & 24 \end{bmatrix} \text{ y } \mathbf{BA} = \begin{bmatrix} -2 & 36 \\ 2 & 20 \end{bmatrix}$$

entonces, por este simple hecho no se puede construir una regla parecida al desarrollo del binomio de Newton

$$\begin{aligned} (\mathbf{A} + \mathbf{B})^2 &= \mathbf{A}^2 + \mathbf{AB} + \mathbf{BA} + \mathbf{B}^2 \\ (\mathbf{A} + \mathbf{B})^3 &= \mathbf{A}^3 + \mathbf{A}^2\mathbf{B} + \mathbf{ABA} + \mathbf{AB}^2 + \mathbf{BA}^2 + \mathbf{BAB} + \mathbf{B}^2\mathbf{A} + \mathbf{B}^3 \end{aligned}$$

y así se complica cada vez más al querer tomar potencias sucesivas. Entonces, si se tienen dos matrices  $\mathbf{A}$  y  $\mathbf{B}$  tales que se cumple que  $\mathbf{AB} = \mathbf{BA}$ , se dice que **conmutan**. Entonces, al conmutar, se tiene un resultado igual al del desarrollo del binomio de Newton con escalares

$$\begin{aligned} (\mathbf{A} + \mathbf{B})^2 &= \mathbf{A}^2 + 2\mathbf{AB} + \mathbf{B}^2 \\ (\mathbf{A} + \mathbf{B})^3 &= \mathbf{A}^3 + 3\mathbf{A}^2\mathbf{B} + 3\mathbf{AB}^2 + \mathbf{B}^3 \end{aligned}$$

Con esta propiedad básica se enuncia el siguiente teorema.

**Teorema 2.4** Sean  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$  tales que conmutan, entonces

$$e^{\mathbf{A}+\mathbf{B}} = e^{\mathbf{A}}e^{\mathbf{B}}$$

*Demostración.* Usando el desarrollo del binomio de Newton

$$\begin{aligned} (\mathbf{A} + \mathbf{B})^n &= \sum_{k=0}^n \binom{n}{k} \mathbf{A}^{n-k} \mathbf{B}^k = \sum_{k=0}^n \frac{n!}{k!(n-k)!} \mathbf{A}^{n-k} \mathbf{B}^k \\ &= \sum_{k=0}^n n! \frac{\mathbf{A}^{n-k}}{(n-k)!} \cdot \frac{\mathbf{B}^k}{k!} \end{aligned}$$

entonces se tiene

$$\frac{(\mathbf{A} + \mathbf{B})^n}{n!} = \sum_{k=0}^n \frac{\mathbf{A}^{n-k}}{(n-k)!} \cdot \frac{\mathbf{B}^k}{k!}$$

tomando la suma infinita

$$\sum_{n=0}^{\infty} \frac{(\mathbf{A} + \mathbf{B})^n}{n!} = \sum_{n=0}^{\infty} \sum_{k=0}^n \frac{\mathbf{A}^{n-k}}{(n-k)!} \cdot \frac{\mathbf{B}^k}{k!}$$

y desarrollando estas sumas se va obteniendo

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{(\mathbf{A} + \mathbf{B})^n}{n!} &= \frac{\mathbf{A}^0 \mathbf{B}^0}{0! 0!} \\ &+ \frac{\mathbf{A}^1 \mathbf{B}^0}{1! 0!} + \frac{\mathbf{A}^0 \mathbf{B}^1}{0! 1!} \\ &+ \frac{\mathbf{A}^2 \mathbf{B}^0}{2! 0!} + \frac{\mathbf{A}^1 \mathbf{B}^1}{1! 1!} + \frac{\mathbf{A}^0 \mathbf{B}^2}{0! 2!} + \dots \end{aligned}$$

por lo que factorizamos las potencias de  $\frac{\mathbf{A}^k}{k!}$

$$\begin{aligned} \sum_{n=0}^{\infty} \frac{(\mathbf{A} + \mathbf{B})^n}{n!} &= \frac{\mathbf{A}^0}{0!} \left( \frac{\mathbf{B}^0}{0!} + \frac{\mathbf{B}^1}{1!} + \frac{\mathbf{B}^2}{2!} + \dots \right) \\ &+ \frac{\mathbf{A}^1}{1!} \left( \frac{\mathbf{B}^0}{0!} + \frac{\mathbf{B}^1}{1!} + \frac{\mathbf{B}^2}{2!} + \dots \right) \\ &+ \frac{\mathbf{A}^2}{2!} \left( \frac{\mathbf{B}^0}{0!} + \frac{\mathbf{B}^1}{1!} + \frac{\mathbf{B}^2}{2!} + \dots \right) + \dots \end{aligned}$$

pero la serie infinita dentro de los paréntesis es igual a  $e^{\mathbf{B}}$ , por lo tanto

$$\begin{aligned} e^{\mathbf{A}+\mathbf{B}} &= \sum_{n=0}^{\infty} \frac{(\mathbf{A} + \mathbf{B})^n}{n!} = \frac{\mathbf{A}^0}{0!} e^{\mathbf{B}} + \frac{\mathbf{A}^1}{1!} e^{\mathbf{B}} + \frac{\mathbf{A}^2}{2!} e^{\mathbf{B}} + \dots \\ &= \left( \frac{\mathbf{A}^0}{0!} + \frac{\mathbf{A}^1}{1!} + \frac{\mathbf{A}^2}{2!} + \dots \right) e^{\mathbf{B}} \\ &= e^{\mathbf{A}} e^{\mathbf{B}} \end{aligned}$$

que es nuestro resultado deseado. ■

**Teorema 2.5** Sea  $\mathbf{A} \in \mathbb{R}^{n \times n}$  una matriz de valores reales. Entonces la matriz exponencial  $\mathbf{F}(t) = e^{\mathbf{A}t}$  satisface la siguiente ecuación diferencial

$$\boxed{\frac{d\mathbf{F}(t)}{dt} = \mathbf{A}\mathbf{F}(t) = \mathbf{F}(t)\mathbf{A}} \tag{2.9}$$

**Demostración.** Para la demostración utilizaremos la definición básica de la derivada

$$\frac{d\mathbf{F}(t)}{dt} = \lim_{h \rightarrow 0} \frac{\mathbf{F}(t+h) - \mathbf{F}(t)}{h}$$

y luego tendremos que utilizar las propiedades básicas de la función exponencial

$$\mathbf{F}(t+h) = e^{\mathbf{A}(t+h)} = e^{\mathbf{A}t} e^{\mathbf{A}h}$$

factorizando obtenemos que

$$\begin{aligned} \mathbf{F}(t+h) - \mathbf{F}(t) &= e^{\mathbf{A}t} e^{\mathbf{A}h} - e^{\mathbf{A}t} \\ &= e^{\mathbf{A}t} (e^{\mathbf{A}h} - \mathbf{I}) \\ &= (e^{\mathbf{A}h} - \mathbf{I}) e^{\mathbf{A}t} \end{aligned}$$

esto por las propiedades básicas de multiplicación de matrices. Entonces nos queda el límite

$$\frac{d\mathbf{F}(t)}{dt} = \lim_{h \rightarrow 0} \frac{e^{\mathbf{A}t} (e^{\mathbf{A}h} - \mathbf{I})}{h} = \lim_{h \rightarrow 0} \frac{(e^{\mathbf{A}h} - \mathbf{I}) e^{\mathbf{A}t}}{h}$$

factorizando lo que no depende de  $h$  obtenemos que

$$\frac{d\mathbf{F}(t)}{dt} = e^{\mathbf{A}t} \left( \lim_{h \rightarrow 0} \frac{e^{\mathbf{A}h} - \mathbf{I}}{h} \right) = \left( \lim_{h \rightarrow 0} \frac{e^{\mathbf{A}h} - \mathbf{I}}{h} \right) e^{\mathbf{A}t}$$

Usando la expansión en series de Taylor de la matriz  $e^{\mathbf{A}h}$  se encuentra a cuanto equivale ese límite.

$$e^{\mathbf{A}h} = \mathbf{I} + \mathbf{A}h + \frac{(\mathbf{A}h)^2}{2!} + \frac{(\mathbf{A}h)^3}{3!} + \dots = \sum_{k=0}^{\infty} \frac{(\mathbf{A}h)^k}{k!}$$

restando  $e^{\mathbf{A}h}$  menos la matriz identidad y reescribiendo obtenemos que

$$e^{\mathbf{A}h} - \mathbf{I} = \mathbf{A}h + \frac{(\mathbf{A}h)^2}{2!} + \frac{(\mathbf{A}h)^3}{3!} + \dots$$

dividiendo entre  $h$  obtenemos

$$\frac{e^{\mathbf{A}h} - \mathbf{I}}{h} = \mathbf{A} + \frac{\mathbf{A}^2 h}{2!} + \frac{\mathbf{A}^3 h^2}{3!} + \dots$$

en Hunter [51] se demuestra que si una sucesión de funciones continuas  $\{f_n\}_{n \in \mathbb{N}}$  con  $f_n : A \rightarrow \mathbb{R}$  converge uniformemente a una función  $f : A \rightarrow \mathbb{R}$ , entonces  $f$  es una función continua en  $A$ , y que esto se puede interpretar como una justificación para poder intercambiar el orden de los límites

$$\lim_{n \rightarrow \infty} \lim_{x \rightarrow a} f_n(x) = \lim_{x \rightarrow a} \lim_{n \rightarrow \infty} f_n(x)$$

y que en nuestro caso nos permite tomar el límite cuando  $h \rightarrow 0$  en cada término de la serie, y los términos que tengan como factor  $h$  desaparecerán

$$\lim_{h \rightarrow 0} \frac{e^{\mathbf{A}h} - \mathbf{I}}{h} = \mathbf{A} + \lim_{h \rightarrow 0} \frac{\mathbf{A}^2 h}{2!} + \lim_{h \rightarrow 0} \frac{\mathbf{A}^3 h^2}{3!} + \dots = \mathbf{A} + \mathbf{0} = \mathbf{A}$$

por lo tanto obtenemos las ecuaciones diferenciales que estabamos buscando

$$\begin{aligned} \frac{d\mathbf{F}(t)}{dt} &= e^{\mathbf{A}t} \left( \lim_{h \rightarrow 0} \frac{e^{\mathbf{A}h} - \mathbf{I}}{h} \right) = \left( \lim_{h \rightarrow 0} \frac{e^{\mathbf{A}h} - \mathbf{I}}{h} \right) e^{\mathbf{A}t} \\ &= e^{\mathbf{A}t} \mathbf{A} = \mathbf{A}e^{\mathbf{A}t} \end{aligned}$$

■

La propiedad anterior es importante dentro de los procesos de Markov, y es el equivalente matricial a  $\frac{d}{dx}[e^{\beta x}] = \beta e^{\beta x}$ . Otras propiedades básicas que nos serán de utilidad sobre la matriz exponencial es poder encontrar las derivadas e integrales de una matriz, y para ello se define a la derivada e integral de una matriz como la derivada e integral elemento a elemento

$$\mathbf{A}(t) = \begin{bmatrix} A_{1,1}(t) & A_{1,2}(t) & \cdots & A_{1,n}(t) \\ A_{2,1}(t) & A_{2,2}(t) & \cdots & A_{2,n}(t) \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1}(t) & A_{n,2}(t) & \cdots & A_{n,n}(t) \end{bmatrix}$$

$$\boxed{\frac{d\mathbf{A}(t)}{dt} = \left[ \frac{dA_{i,j}(t)}{dt} \right]_{i,j} \quad \int \mathbf{A}(t)dt = \left[ \int A_{i,j}(t)dt \right]_{i,j}} \quad (2.10)$$

donde las funciones  $A_{i,j}(t)$  con  $i, j = 1, 2, \dots, n$  se asumen que son continuas, derivables e integrables para cada valor  $t \in \mathbb{R}$ . Como ejemplo básico de esta definición se muestra el siguiente ejemplo práctico

$$\mathbf{A}(x) = \begin{bmatrix} x^2 & e^{-x} \\ \cos(x) & -\sin(x) \end{bmatrix}$$

$$\text{a) } \frac{d\mathbf{A}(x)}{dx} = \begin{bmatrix} 2x & -e^{-x} \\ -\sin(x) & -\cos(x) \end{bmatrix} \quad \text{b) } \int \mathbf{A}(x)dx = \begin{bmatrix} \frac{1}{3}x^3 & -e^{-x} \\ \sin x & \cos x \end{bmatrix}$$

Y estas definiciones serán de mucha utilidad al querer encontrar propiedades básicas donde sea necesario calcular la derivada/integral de una matriz. Otras propiedades básicas de la matriz exponencial se muestran a continuación.

**Teorema 2.6** Sean  $\mathbf{F}(t), \mathbf{G}(t) : D \subset \mathbb{R} \rightarrow \mathbb{R}^{n \times n}$  funciones con valores en  $\mathbb{R}^{n \times n}$  continuas y derivables. Entonces se cumple que

$$\boxed{\frac{d}{dt} [\mathbf{F}(t)\mathbf{G}(t)] = \frac{d\mathbf{F}(t)}{dt} \mathbf{G}(t) + \mathbf{F}(t) \frac{d\mathbf{G}(t)}{dt}} \quad (2.11)$$

Con este resultado se puede deducir fácilmente este otro resultado básico equivalente al de derivar una función de una variable

$$\boxed{\frac{d}{dt} [\mathbf{A}\mathbf{B}(t)] = \mathbf{A} \frac{d}{dt} [\mathbf{B}(t)]} \quad (2.12)$$

En palabras nos dice que si tenemos una matriz de constantes multiplicada por una matriz funcional, entonces la derivada del producto es igual al producto la matriz de constantes por la derivada de la matriz funcional, "sacamos" la matriz de constantes del operador derivada. También con esta definición de matrices de funciones podemos obtener un corolario que nos será de mucha utilidad más adelante cuando deseemos obtener integrales de matrices con estructuras más elaboradas.

**Corolario 2.7** *Sea  $\mathbf{A} \in \mathbb{R}^{n \times n}$  una matriz de constantes y sea  $\mathbf{B}(x) \in \mathbb{R}^{n \times n}$  una matriz que dependa de un parámetro o variable  $x$ , y  $\beta$  una constante cualquiera. Entonces se cumplen las siguientes identidades:*

$$\boxed{\begin{aligned} a) \int [\mathbf{A} + \beta\mathbf{B}(x)] dx &= \mathbf{A}x + \beta \int \mathbf{B}(x)dx \\ b) \int \beta\mathbf{A}\mathbf{B}(x)dx &= \beta\mathbf{A} \int \mathbf{B}(x)dx \end{aligned}} \quad (2.13)$$

**Demostración.** *Estas propiedades se pueden obtener fácilmente mediante la aplicación de las reglas del cálculo. Para la primera identidad obtenemos*

$$\int [\mathbf{A} + \beta\mathbf{B}(x)] dx = \int \mathbf{A}dx + \int \beta\mathbf{B}(x)dx = \mathbf{A} \int dx + \beta \int \mathbf{B}(x)dx = \mathbf{A}x + \beta \int \mathbf{B}(x)dx$$

*y para la segunda propiedad tenemos que utilizar la definición formal de multiplicación de matrices. Sea  $\mathbf{C} = \beta\mathbf{A}\mathbf{B}(x)$  entonces obtenemos que*

$$C_{i,j} = \sum_{k=1}^n \beta A_{i,k} B_{k,j}(x)$$

*y por lo tanto la integral de la entrada  $C_{i,j}$  está dada por*

$$\int C_{i,j} dx = \int \sum_{k=1}^n \beta A_{i,k} B_{k,j}(x) dx = \sum_{k=1}^n \int \beta A_{i,k} B_{k,j}(x) dx = \sum_{k=1}^n \beta A_{i,k} \int B_{k,j}(x) dx$$

*Una nota importante es que el intercambio de la suma y de la integral es válido en el caso de una suma finita de elementos. En términos matriciales generales esto implica que*

$$\int \beta\mathbf{A}\mathbf{B}(x) dx = \beta\mathbf{A} \int \mathbf{B}(x) dx$$

*que es lo que queríamos mostrar. En palabras nos dice que podemos sacar matrices de coeficientes constantes de una integral. ■*

Juntando el resultado anterior (2.13) junto con el de la derivada de la matriz exponencial (2.9) se obtiene el equivalente de la integral de la matriz exponencial.

**Teorema 2.8** Si  $\mathbf{A} \in \mathbb{R}^{n \times n}$  es una matriz no singular y  $\mathbf{C} \in \mathbb{R}^{n \times n}$  una matriz arbitraria dada entonces

$$\int e^{\mathbf{A}t} dt = \mathbf{A}^{-1} e^{\mathbf{A}t} + \mathbf{C} \quad (2.14)$$

*Demostración.* Para ello recordemos primeramente que

$$\frac{d}{dt}[e^{\mathbf{A}t}] = \mathbf{A}e^{\mathbf{A}t}$$

y del resultado (2.12)

$$\frac{d}{dt} [\mathbf{A}^{-1}e^{\mathbf{A}t} + \mathbf{C}] = \mathbf{A}^{-1} \frac{d}{dt}[e^{\mathbf{A}t}] + \frac{d}{dt}[\mathbf{C}] = \mathbf{A}^{-1} [\mathbf{A}e^{\mathbf{A}t}] = e^{\mathbf{A}t}$$

integrando de ambos lados y usando el teorema fundamental del cálculo nos queda

$$\mathbf{A}^{-1}e^{\mathbf{A}t} + \mathbf{C} = \int e^{\mathbf{A}t} dt$$

que es lo que se quería demostrar. ■

## 2.2 Método de la diagonalización

En el siguiente capítulo se analizará la estructura matemática de las matrices que surgen al trabajar con los procesos de Markov a tiempo continuo. En los procesos de Markov surge una matriz  $\mathbf{T}$  que se denota como la "matriz de tasas"<sup>2</sup>. Inicialmente solamente se revisarán algunos métodos básicos para encontrar la exponencial de la matriz  $\mathbf{T} \in \mathbb{R}^{n \times n}$  dada por  $e^{\mathbf{T}}$ . Esta matriz de tasas tiene la propiedad de que los elementos fuera de la diagonal son no negativos y cada elemento de la diagonal es el inverso aditivo de la suma por renglón.

$$T_{i,j} \geq 0 \text{ para } i \neq j \text{ y } T_{i,i} = - \sum_{i \neq j} T_{i,j}$$

y por lo tanto esta matriz  $\mathbf{T}$  su suma por renglones es cero. Una primera interpretación de esta matriz de tasas es que muestra como se "comunican" cada pareja de estados  $i, j$  con  $i \neq j$ . A partir de esta matriz de tasas se desea encontrar las probabilidades de transición entre estados para cada  $t \in \mathbb{R}^+$ , que matricialmente se le denota por  $\mathbf{P}(t) \in \mathbb{R}^{n \times n}$  y esta matriz de probabilidades cumple que sus entradas son no negativas y que la suma por renglones es uno

$$\mathbf{P}(t) \geq \mathbf{0} \text{ y } \mathbf{P}(t)\mathbf{e}' = \mathbf{e}' \text{ para } t \in \mathbb{R}^+$$

---

<sup>2</sup>En la literatura se le denota comúnmente a esta matriz  $\mathbf{T}$  como  $\mathbf{Q}$  y la llaman "matriz generadora infinitesimal". En este trabajo se le denotará de manera más práctica como "matriz de tasas" y por eso es que se le denotará por  $\mathbf{T}$ .

Para encontrar estas probabilidades de transición, es necesario resolver el siguiente sistema de ecuaciones diferenciales

$$\frac{d\mathbf{P}(t)}{dt} = \mathbf{T}\mathbf{P}(t) = \mathbf{P}(t)\mathbf{T} \quad \text{con } \mathbf{P}(0) = \mathbf{I}. \quad (2.15)$$

llamadas las "**ecuaciones hacia atrás/adelante**" de Kolmogorov. Sabemos por (2.9) que estas ecuaciones tienen como solución exacta

$$\mathbf{P}(t) = e^{\mathbf{T}t} \quad \text{para } t \in [0, \infty) \quad (2.16)$$

y que una forma de calcular estas probabilidades es calculando la serie infinita

$$\begin{aligned} \mathbf{P}(t) &= \mathbf{I} + \mathbf{T}t + \frac{(\mathbf{T}t)^2}{2} + \frac{(\mathbf{T}t)^3}{3!} + \dots \\ &= \mathbf{I} + \mathbf{T}t + \frac{\mathbf{T}^2 t^2}{2} + \frac{\mathbf{T}^3 t^3}{3!} + \dots \end{aligned}$$

donde existe el inconveniente de que no sabremos de antemano cuántos sumandos hay que tomar para obtener una buena aproximación a la probabilidad buscada. El problema al que nos enfrentamos con esta definición es que no se podría realizar esta suma infinita y se tienen que realizar trucamientos que garanticen exactitud. Otro problema es que se tendría que realizar en varias ocasiones estas sumas si es que uno desea encontrar las probabilidades en varios momentos del tiempo  $t_1, t_2, \dots, t_m$ , y analizar la *historia* de como son las probabilidad de transición para cada uno de los estados. Un inconveniente más al utilizar esta definición es que se debe calcular potencias sucesivas de la matriz de tasas  $\mathbf{T}$  y computacionalmente sería un proceso muy laborioso. Para resolver este problema usaremos los siguientes resultados básicos del álgebra lineal.

**Definición 2.9** Sean  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$ . Se dice que  $\mathbf{A}$  es **semejante** a  $\mathbf{B}$  si existe una matriz  $\mathbf{P} \in \mathbb{R}^{n \times n}$  tal que

$$\mathbf{AP} = \mathbf{PB} \quad (2.17)$$

Si esta matriz  $\mathbf{P}$  de la definición anterior es invertible, entonces se puede escribir como

$$\mathbf{A} = \mathbf{PBP}^{-1}$$

En Strang [49] se muestra cuáles son las hipótesis para que una matriz sea diagonalizable.

**Definición 2.10** Una matriz  $\mathbf{A} \in \mathbb{R}^{n \times n}$  es **diagonalizable** si existe una matriz  $\mathbf{D}$  cuyos elementos que no están en la diagonal son iguales a cero y  $\mathbf{A}$  es semejante a  $\mathbf{D}$ .

**Teorema 2.11** Una matriz  $\mathbf{A} \in \mathbb{R}^{n \times n}$  es **diagonalizable** si tiene  $n$  vectores propios linealmente independientes. Si estos vectores característicos son las columnas de una matriz  $\mathbf{P}$  entonces la matriz  $\mathbf{D}$  es una matriz diagonal

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} \quad (2.18)$$

donde  $\mathbf{P} = [\mathbf{v}'_1 \ \mathbf{v}'_2 \ \cdots \ \mathbf{v}'_n]$  y  $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$  donde  $\mathbf{v}'_k$  y  $\lambda_k$  son los vectores propios y valores propios asociados respectivamente.

Dos observaciones que se hacen es que la diagonalización de  $\mathbf{A}$  depende de la existencia de suficientes vectores característicos y que la invertibilidad de  $\mathbf{A}$  depende de los valores característicos distintos de cero. Después, para matrices semejantes, existen dos resultados básicos sobre la matriz exponencial

**Teorema 2.12** Si  $\mathbf{D} \in \mathbb{R}^{n \times n}$  es una matriz diagonal entonces

$$e^{\mathbf{D}} = \text{diag}(e^{d_{11}}, e^{d_{22}}, \dots, e^{d_{nn}}) \quad (2.19)$$

**Demostración.** La parte importante de recalcar es que si se tiene una matriz diagonal, entonces potencias sucesivas de la matriz se calculan encontrando las potencias de cada entrada de la diagonal

$$\begin{aligned} \mathbf{D} &= \text{diag}(d_{11}, d_{22}, \dots, d_{n,n}) \\ \mathbf{D}^k &= \text{diag}(d_{11}^k, d_{22}^k, \dots, d_{n,n}^k) \quad \text{con } k \geq 1 \end{aligned}$$

entonces, al calcular la serie infinita

$$\begin{aligned} e^{\mathbf{D}} &= \mathbf{I} + \mathbf{D} + \frac{\mathbf{D}^2}{2} + \frac{\mathbf{D}^3}{3!} + \dots = \text{diag}\left(\sum_{k=0}^{\infty} \frac{d_{11}^k}{k!}, \sum_{k=0}^{\infty} \frac{d_{22}^k}{k!}, \dots, \sum_{k=0}^{\infty} \frac{d_{n,n}^k}{k!}\right) \\ &= \begin{bmatrix} e^{d_{11}} & 0 & \dots & 0 \\ 0 & e^{d_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{d_{nn}} \end{bmatrix} \end{aligned}$$

que es lo que se quería demostrar. ■

**Teorema 2.13** Sean  $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}$  matrices semejantes. Entonces se tiene que

$$e^{\mathbf{A}} = \mathbf{P}e^{\mathbf{B}}\mathbf{P}^{-1} \quad (2.20)$$

**Demostración.** Solamente para demostrar este hecho se tiene que tomar en cuenta que

$$(\mathbf{PBP}^{-1})^k = \mathbf{PB}^k\mathbf{P}^{-1} \text{ para } k \geq 1$$

por lo tanto

$$\begin{aligned} e^{\mathbf{A}} &= \sum_{k=0}^{\infty} \frac{(\mathbf{PBP}^{-1})^k}{k!} = \sum_{k=0}^{\infty} \frac{\mathbf{PB}^k\mathbf{P}^{-1}}{k!} \\ &= \mathbf{P} \left( \sum_{k=0}^{\infty} \frac{\mathbf{B}^k}{k!} \right) \mathbf{P}^{-1} = \mathbf{P}e^{\mathbf{B}}\mathbf{P}^{-1} \end{aligned}$$

y es el resultado buscado. ■

Con estos resultados preliminares se obtiene el siguiente resultado, que nos permitirá construir de manera fácil las probabilidades de transición de un proceso markoviano con matriz de tasas  $\mathbf{T}$

**Teorema 2.14** Sea  $\{X(t)\}_{t \in \mathbb{R}^+}$  un proceso estocástico markoviano con  $n$  estados y matriz de tasas  $\mathbf{T} \in \mathbb{R}^{n \times n}$  diagonalizable asociado a (2.15). Entonces las probabilidades de transición están dadas por

$$\mathbf{P}(t) = \mathbf{V}e^{\mathbf{D}t}\mathbf{V}^{-1} \text{ con } t \in \mathbb{R}^+ \quad (2.21)$$

donde  $\mathbf{T} = \mathbf{VDV}^{-1}$  y las columnas de la matriz  $\mathbf{V}$  son los vectores propios de  $\mathbf{T}$  y  $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$  la matriz diagonal con los valores propios.

**Demostración.** En la teoría de procesos de Markov existe una relación fundamental que se conoce como las **ecuaciones de Chapman-Kolmogorov**, que relacionan las probabilidades de transición en diferentes etapas del tiempo

$$\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s)$$

para cada pareja  $t, s \in \mathbb{R}^+$ . En Rolski et. al [61] se demuestra que la matriz de tasas  $\mathbf{T}$  cumple la siguiente relación como "matriz de intensidades"

$$\mathbf{T} = \left. \frac{d\mathbf{P}(t)}{dt} \right|_{t=0} = \lim_{h \rightarrow 0^+} \frac{\mathbf{P}(h) - \mathbf{I}}{h} \quad (2.22)$$

y con esta relación se obtienen las ecuaciones de Kolmogorov hacia adelante/atrás

$$\begin{aligned} \frac{d\mathbf{P}(t)}{dt} &= \lim_{h \rightarrow 0^+} \frac{\mathbf{P}(t+h) - \mathbf{P}(t)}{h} \\ &= \lim_{h \rightarrow 0^+} \frac{\mathbf{P}(t)\mathbf{P}(h) - \mathbf{P}(t)}{h} = \lim_{h \rightarrow 0^+} \frac{\mathbf{P}(h)\mathbf{P}(t) - \mathbf{P}(t)}{h} \\ &= \mathbf{P}(t) \lim_{h \rightarrow 0^+} \frac{\mathbf{P}(h) - \mathbf{I}}{h} = \lim_{h \rightarrow 0^+} \frac{\mathbf{P}(h) - \mathbf{I}}{h} \mathbf{P}(t) \\ &= \mathbf{P}(t)\mathbf{T} = \mathbf{TP}(t) \end{aligned}$$



```

for q=1:n(3)
    vector(q)=A(i,j,q);
end

```

Esta función extrae cada renglón de esta malla, ya con ello estamos en condiciones de poder calcular las probabilidades de transición de procesos cuyas matrices de tasas sean diagonalizables. Un problema numérico es que el cálculo de los valores y vectores propios, ya que entre más grande es la matriz, entonces este procedimiento se vuelve laborioso, con la dificultad de que el rango de la matriz  $\mathbf{V}$  ya no sea completo, y con esto la imposibilidad de calcular la inversa  $\mathbf{V}^{-1}$ . En Meyer [1] se explica que una matriz es **mal condicionada** si un pequeño cambio en la matriz  $\mathbf{A}$  ocasiona un cambio relativo largo en la inversa  $\mathbf{A}^{-1}$ , y que este grado de mal condicionamiento se puede medir encontrando el **número de condición** de una matriz, dado por  $\kappa(\mathbf{A}) = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|$ , con  $\|\cdot\|$  una norma matricial cualquiera. En Cheney et. al. [68] se muestra una regla básica que si la magnitud del número de condición es de  $10^p$ , entonces uno puede esperar de perder al menos  $p$  dígitos de precisión al resolver el sistema de ecuaciones  $\mathbf{A}\mathbf{x}' = \mathbf{b}'$ , y que esto se aplicaría al encontrar la inversa de la matriz. Y en el análisis del problema de diagonalización, si la matriz de vectores propios es bien condicionada, se espera una magnitud de  $\kappa(\mathbf{V})$  relativamente menor y el residual<sup>3</sup>  $\|\mathbf{V}\mathbf{V}^{-1} - \mathbf{I}\|$  pequeño.

```

function Pt = diagonalizacion(T,tau)

% ESTA FUNCION CALCULA LAS PROBABILIDADES DE TRANSICION DEL
% PROCESO ESTOCASTICO CON MATRIZ DE TASAS "T" HASTA EL TIEMPO
% "tau".
%
% FORMA DE USARSE: Pt = diagonalizacion(T,tau)

[n1,n2]=size(T);
if(n1~=n2)
    error(' LA MATRIZ DEBE SER CUADRADA!!! ')
end

n=n1;

% Construimos la particion del tiempo
xt = 0:0.1:tau;
m = length(xt);

% Calculamos los valores y vectores propios de la matriz de tasas

```

<sup>3</sup>En Matlab, la unidad de redondeo (o la precisión de la máquina) se denota por la letra  $u$  y que aproximadamente es del orden  $10^{-8}$  para computadores de precisión simple y de  $10^{-16}$  para computadores de doble precisión. De manera práctica se puede pensar que la unidad de redondeo es aquel número tal que  $1 + u \neq 1$ . En Higham [10] se pueden ver más detalles sobre la unidad de redondeo y sus propiedades.

```

[V,D]=eig(T);

% Calculamos las probabilidades de transicion P(t) = V e^(Dt) inv(V)
Pt = zeros(n,n,m); invV = inv(V);

% Se calcula el numero de condicion y el residual
K2= cond(V); Residual = norm(V*invV - eye(n),2);

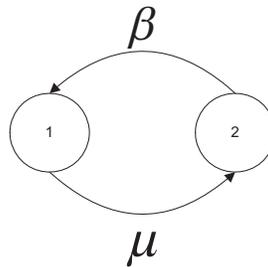
for k=1:m
    Pt(:,:,k) = V*diag(exp(xt(k)*diag(D)))*invV;
end

% Por ultimo, graficamos estas probabilidades para cada (i,j)
cont=1;
for i = 1:n
    for j = 1:n
        vecPij=copia_valores_multidimensional(Pt,i,j);
        subplot(n,n,cont), plot(xt,vecPij,'LineWidth',2.2)
        title(sprintf(' Pt( %d , %d ) ', i,j ))
        cont=cont+1;
    end
end

disp(sprintf(' Num Cond = %g , Norma residual = %g ', K2, Residual))

```

El caso más práctico donde se puede verificar que la matriz de tasas es diagonalizable es cuando tenemos dos estados  $\mathbb{E} = \{1, 2\}$  y las transiciones están dadas por



Más adelante se analizará el porqué que la matriz de tasas para este proceso está dado por

$$\mathbf{T} = \begin{bmatrix} -\mu & \mu \\ \beta & -\beta \end{bmatrix}.$$

La característica más importante de las tasas de transición es que debido a como esta construido la matriz de tasas se puede observar que el valor  $\lambda = 0$  es un valor propio y el vector unitario  $\mathbf{e}'$  es el vector propio asociado.

$$\mathbf{T}\mathbf{e}' = \mathbf{0}' \implies \mathbf{T}\mathbf{e}' = 0\mathbf{e}'.$$

para encontrar el otro valor propio primero encontramos el polinomio característico de la matriz  $p(\lambda) = \det(\mathbf{T} - \lambda\mathbf{I})$

$$\begin{aligned} \det \left( \begin{bmatrix} -\mu - \lambda & \mu \\ \beta & -\beta - \lambda \end{bmatrix} \right) &= (-\mu - \lambda)(-\beta - \lambda) - \beta\mu \\ &= \lambda^2 + \lambda(\mu + \beta) \end{aligned}$$

y resolviendo la ecuación  $p(\lambda) = 0$  se obtiene

$$\begin{aligned} \lambda^2 + \lambda(\mu + \beta) &= 0 \\ \lambda = 0 \text{ ó } \lambda + \mu + \beta &= 0 \end{aligned}$$

$$\lambda_1 = 0 \text{ ó } \lambda_2 = -\mu - \beta$$

Ya tenemos cuál es el vector propio asociado a  $\lambda_1 = 0$ , solo nos falta encontrar el vector propio para  $\lambda_2 = -\mu - \beta$ . Queremos encontrar  $\mathbf{v}_2 = \begin{bmatrix} x & y \end{bmatrix}$  tal que

$$\begin{bmatrix} -\mu - (-\mu - \beta) & \mu \\ \beta & -\beta - (-\mu - \beta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

y nos queda

$$\begin{bmatrix} \beta & \mu \\ \beta & \mu \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \implies \begin{aligned} \beta x + \mu y &= 0 \\ \beta x + \mu y &= 0 \end{aligned}$$

con esto podemos ver fácilmente que  $x = 1$  y  $y = -\mu/\beta$  son los elementos del segundo vector propio, por lo tanto podemos encontrar fácilmente las probabilidades de transición  $\mathbf{P}(t)$  están dadas por

$$\begin{aligned} \mathbf{P}(t) &= \begin{bmatrix} 1 & 1 \\ 1 & -\mu/\beta \end{bmatrix} \begin{bmatrix} e^{0t} & 0 \\ 0 & e^{-(\mu+\beta)t} \end{bmatrix} \begin{bmatrix} \frac{\beta}{\beta + \mu} & \frac{\mu}{\beta + \mu} \\ \frac{\mu}{\beta + \mu} & -\frac{\mu}{\beta + \mu} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\beta}{\beta + \mu} + \frac{\mu e^{-(\mu+\beta)t}}{\beta + \mu} & \frac{\mu}{\beta + \mu} - \frac{\mu e^{-(\mu+\beta)t}}{\beta + \mu} \\ \frac{\beta}{\beta + \mu} - \frac{\beta e^{-(\mu+\beta)t}}{\beta + \mu} & \frac{\mu}{\beta + \mu} + \frac{\beta e^{-(\mu+\beta)t}}{\beta + \mu} \end{bmatrix} \end{aligned}$$

Para analizar numéricamente este procedimiento, usaremos la matriz de tasas

$$\mathbf{T} = \begin{bmatrix} -2 & 2 \\ 0.01 & -0.01 \end{bmatrix}$$

donde se utilizan los valores  $\mu = 2$  y  $\beta = 0.01$

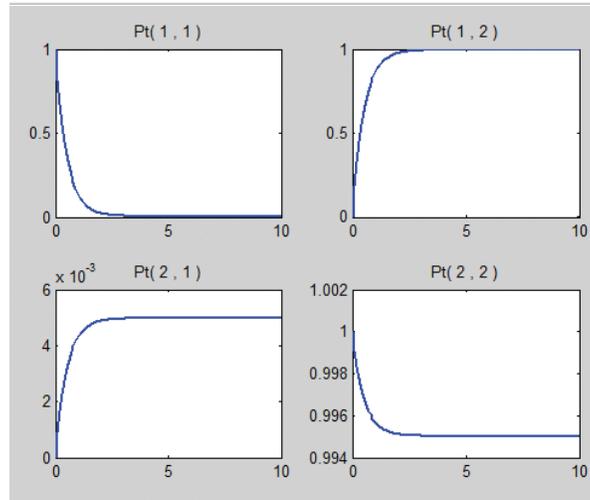
```
>> T=[-2 2; 0.01 -0.01]
```

```
T =
```

```
-2.0000    2.0000
 0.0100   -0.0100
```

```
>> Pt = diagonalizacion(T,10);
```

```
Num Cond = 2.39725 , Norma residual = 1.57009e-016
```



Recordemos que las probabilidades  $P_{i,j}(t)$  en términos de probabilidad condicional están dadas por  $P[X(t) = j | X(0) = i]$ , lo que en nuestro caso  $P_{1,2}(t)$  significa que si empezamos en el estado 1 la probabilidad de que al tiempo  $t = 5$  este en el estado 2 es muy cercano a 1, por lo que esperaremos ver esta transición de manera rápida. Es importante señalar que el trabajo computacional más laborioso es encontrar los valores y vectores propios de la matriz aunque que para matrices de regular tamaño actualmente no es un problema muy complicado de realizar. Una recomendación es hacer un análisis previo de los valores propios y vectores propios ya que el método de la diagonalización depende de la exactitud con la que se obtiene la matriz  $\mathbf{D}$  y  $\mathbf{V}$ . Regresando al ejemplo anterior también nos muestra que las probabilidades se estabilizan en un cierto momento, y esto es cierto cuando la matriz de tasas representa a un proceso irreducible, por lo que la matriz límite existe y se denota por  $\mathbf{\Pi}$

$$\mathbf{\Pi} = \lim_{t \rightarrow \infty} \mathbf{P}(t)$$

en nuestro ejemplo esta matriz está dada

$$\mathbf{\Pi} = \begin{bmatrix} 0.005 & 0.995 \\ 0.005 & 0.995 \end{bmatrix}$$

esto significa que el proceso estocástico visita a la larga el estado 2 el 99.5% de las veces y que el estado 1 sólo se visita el 0.5% de las veces, y por ello es importante en la práctica poder encontrar esta matriz límite  $\mathbf{\Pi}$ . Más adelante se verán métodos que garanticen el cálculo eficiente de esta matriz límite.

### 2.3 Método de la uniformización

Hemos visto que si una matriz de tasas  $\mathbf{T}$  es diagonalizable, entonces se pueden obtener fácilmente las probabilidades de transición  $\mathbf{P}(t)$  dadas por

$$\mathbf{P}(t) = \mathbf{V}e^{\mathbf{D}t}\mathbf{V}^{-1}.$$

donde  $\mathbf{V}$  es la matriz que está formada por los vectores propios de la matriz  $\mathbf{T}$  y la matriz  $\mathbf{D}$  es una matriz diagonal formada por los valores propios de  $\mathbf{T}$ . El cálculo de los valores y vectores propios puede ser un problema numéricamente inestable y si en el peor de los casos no se tiene que todos los valores propios sean distintos, entonces este resultado  $\mathbf{P}(t) = \mathbf{V}e^{\mathbf{D}t}\mathbf{V}^{-1}$  pierde su importancia práctica, ya que se tendría que recurrir al cálculo de la forma canónica de Jordan de una matriz, y en Golub-Wilkinson [43] se demuestra que el cálculo de la forma de Jordan es un problema numéricamente inestable. Para tratar con el problema de encontrar las probabilidades de transición de un proceso general y evitar el cálculo de las matrices  $\mathbf{T}^n$ , se utiliza el siguiente método popular en la literatura. Se construye una matriz de transición  $\tilde{\mathbf{P}}$  dada por

$$\tilde{\mathbf{P}} = \frac{\mathbf{T}}{\lambda} + \mathbf{I}$$

donde el valor  $\lambda$  es tal que cumple  $\lambda \geq \max |T_{i,i}|$ , el valor más grande en valor absoluto de la diagonal de la matriz de tasas. Entonces, si despejamos a la matriz  $\mathbf{T}$  de la ecuación anterior obtenemos que

$$\mathbf{T} = \lambda (\tilde{\mathbf{P}} - \mathbf{I})$$

y por lo tanto

$$\begin{aligned} \mathbf{P}(t) &= e^{\mathbf{T}t} \\ &= e^{\lambda(\tilde{\mathbf{P}}-\mathbf{I})t} \\ &= e^{\lambda\tilde{\mathbf{P}}t-\lambda\mathbf{I}t} \\ &= e^{\lambda\tilde{\mathbf{P}}t}e^{-\lambda\mathbf{I}t} \\ &= e^{\lambda\tilde{\mathbf{P}}t}e^{-\lambda t}\mathbf{I} \\ &= \sum_{k=0}^{\infty} \frac{(\lambda t)^k \tilde{\mathbf{P}}^k}{k!} \cdot e^{-\lambda t} \\ &= \sum_{k=0}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \tilde{\mathbf{P}}^k. \end{aligned}$$

Estas relaciones son ciertas debido a que las matrices conmutan. Después, al analizar el primer factor se identifica que es una distribución de Poisson con parámetro  $\lambda t$  y la matriz  $\tilde{\mathbf{P}}^k$  representa las probabilidades de transición de algún proceso estocástico discreto. Una de las ventajas que se tiene para usar el método de uniformización es que se puede elegir un valor entero  $N$ , que dependerá de un error de tolerancia  $\varepsilon$  que nosotros elegimos, tal que

$$\begin{aligned} \|\mathbf{P}(t) - \mathbf{P}_N(t)\| &= \left\| \sum_{k=0}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \tilde{\mathbf{P}}^k - \sum_{k=0}^N \frac{(\lambda t)^k e^{-\lambda t}}{k!} \tilde{\mathbf{P}}^k \right\| \\ &= \left\| \sum_{k=N+1}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \tilde{\mathbf{P}}^k \right\| \leq \varepsilon \end{aligned}$$

usando las propiedades de las normas matriciales

$$\begin{aligned}
\left\| \sum_{k=N+1}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \tilde{\mathbf{P}}^k \right\| &\leq \sum_{k=N+1}^{\infty} \left\| \frac{(\lambda t)^k e^{-\lambda t}}{k!} \tilde{\mathbf{P}}^k \right\| \\
&\leq \sum_{k=N+1}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \|\tilde{\mathbf{P}}^k\| \\
&\leq \sum_{k=N+1}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \\
&\leq \varepsilon
\end{aligned} \tag{2.23}$$

Esto se debe a que para las matrices de transición su norma infinito es uno  $\|\mathbf{P}\|_{\infty} = 1$ <sup>4</sup>, y la suma  $\sum_{k=N+1}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!}$  es la probabilidad de que al menos ocurran  $N$  eventos en una distribución Poisson de parámetro  $\lambda t$ , por lo que

$$\begin{aligned}
P[\text{Poisson} > N] &\leq \varepsilon \\
1 - P[\text{Poisson} \leq N] &\leq \varepsilon \\
P[\text{Poisson} \leq N] &\geq 1 - \varepsilon.
\end{aligned} \tag{2.24}$$

Con esto, se buscará ese valor  $N$  especial que satisfaga que su función de distribución cumpla esta desigualdad (2.24), por lo que estaremos siempre asumiendo que existirá un error en el cálculo de nuestras probabilidades. No se recomienda que el error sea muy pequeño porque si no tendremos que calcular sumas con demasiados términos y por lo tanto computacionalmente se volverá intratable. La estrategia un poco rudimentaria será la siguiente: Elegimos un valor  $t = \tau$  "grande", seleccionamos un error  $\varepsilon^*$ , y encontramos el valor "óptimo"  $N_{opt}$ , si las matrices no son muy grandes, podremos almacenar en memoria todas las matrices  $\tilde{\mathbf{P}}^k$  con  $k = 0, 1, 2, \dots, N_{opt}$  en un arreglo multidimensional, para después ir calculando las probabilidades en tiempos  $0 < t_1 < t_2 < \dots \leq t_n = \tau$  y poder graficar los resultados obtenidos. La idea central para estimar las probabilidades  $P_{i,j}(t_k)$  es construir este arreglo

---

<sup>4</sup>Hay dos normas matriciales importantes, la norma-1 y la norma infinito

$$\begin{aligned}
\|\mathbf{A}\|_1 &= \max_j \sum_i |A_{ij}| \\
\|\mathbf{A}\|_{\infty} &= \max_i \sum_j |A_{ij}|
\end{aligned}$$

que denotan respectivamente el máximo de las sumas por columna y el máximo de la suma por renglones respectivamente. Para el caso de la desigualdad (2.23) se utilizó las propiedades de la norma infinito debido a que las matrices de transición su suma por renglón es 1.



```

% Construimos la particion del tiempo
xt = 0:0.1:tau; m = length(xt);

% Encontramos el valor Nopt de la distribucion Poisson
Nopt = poissinv(1-err,tau*lam);

% Generamos un vector de apoyo para el calculo posterior de
% las probabilidades Poisson
xapo = 0:Nopt;

% Generamos "la malla" de matrices de transicion P^k
Pprob = zeros(n,n,Nopt+1); Pprob(:,:,1) = eye(n);

for k=2:Nopt+1
    Pprob(:,:,k) = Pest*Pprob(:,:,k-1);
end

% Generamos "la malla" de la probabilidad P(t) para cada 0<t1<t2<...<tau
summat = zeros(n); Pt = zeros(n,n,m);

for k = 1:m
    Poisson = poisspdf(xapo, lam*xt(k));
    % Generamos las probabilidades Poisson
    for q = 1:Nopt+1
        summat = summat + Poisson(q)*Pprob(:,:,q);
    end
    Pt(:,:,k) = summat; summat = zeros(n);
end

% Graficamos las probabilidades de transicion
cont=1;

for i = 1:n
    for j = 1:n
        vecPij=copia_valores_multidimensional(Pt,i,j);
        subplot(n,n,cont), plot(xt,vecPij), title(sprintf(' Pt( %d , %d ) ', i,j ))
        cont=cont+1;
    end
end
end

```

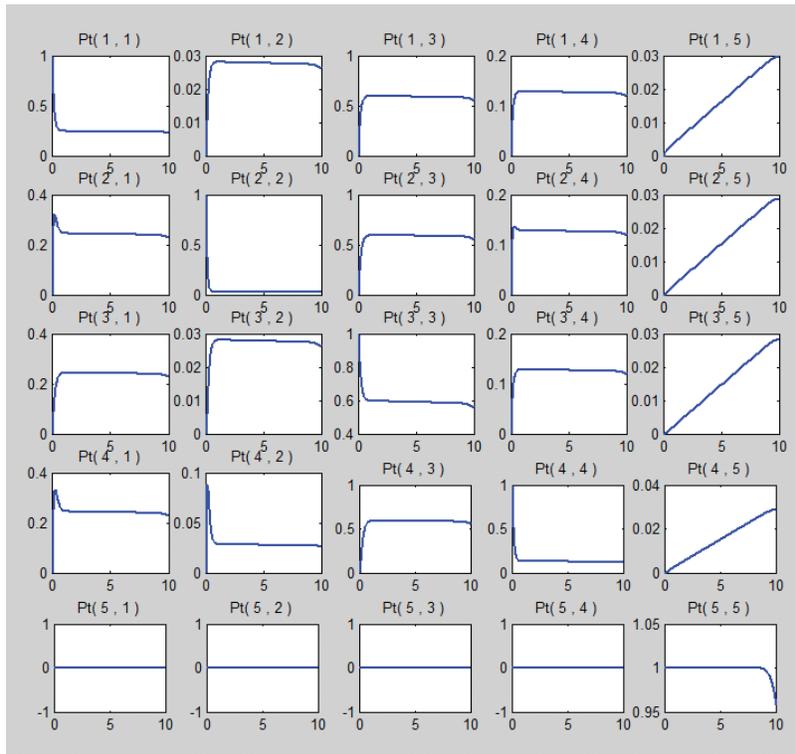
Un ejemplo práctico de este resultado y programa se muestra a continuación. Supongamos que tenemos un proceso de Markov  $\{X(t)\}_{t \in \mathbb{R}^+}$  con espacio de estados  $\mathbb{E} = \{1, 2, 3, 4, 5\}$  y

matriz de tasas de intensidad dada por

$$\mathbf{T} = \begin{bmatrix} -5.11 & 0.1 & 4 & 1 & 0.01 \\ 5 & -10 & 3 & 2 & 0 \\ 1 & 0 & -2.001 & 1 & 0.001 \\ 4 & 2 & 1 & -7 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (2.25)$$

donde el estado 5 es un estado absorbente (porque al entrar a él no se podrá visitar nunca más otro estado), y los demás son transitorios. Calculando las probabilidades de transición del proceso hasta el tiempo  $\tau = 10$  y con un error de tolerancia  $\varepsilon = 0.05$  obtenemos los siguientes resultados gráficos.

```
>> T=[-5.11 0.1 4 1 0.01; 5 -10 3 2 0; 1 0 -2.001 1 0.001;
      4 2 1 -7 0; 0 0 0 0 0];
>> Pt = uniformizacion(T,10,0.05);
```



Esto nos dice que la probabilidad de ser absorbido al estado 5 es creciente para  $P_{i,5}(t)$ , aunque aquí tenemos el problema de no saber exactamente en que momento uno entrará a ese estado, por lo que será necesario calcular la esperanza del tiempo esperado de primer arribo

$$A = \inf \{t > 0 \mid X(t) = 5\} .$$

más adelante veremos al calcular las probabilidades de absorción. En la práctica, este primer código puede resultar ser de muy poca utilidad, debido a la cantidad de cálculos internos que se están realizando para ir estimando las probabilidades  $\mathbf{P}(t_k)$ . Una forma práctica de

ahorrarse un poco de trabajo computacional es utilizar las propiedades de la ecuación de Chapman-Kolmogorov  $\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s)$  y con esto solamente tenemos que calcular con una mejor precisión el primer tiempo (que en nuestro caso hemos fijado como  $t = 0.1$ ) y después ir realizando las sucesivas multiplicaciones de matrices correspondientes. El código lo llamaremos "uniformizacion2.m", y solo se mostrará la parte del cambio central, ya que la mayoría del código se repite.

```
function Pt = uniformizacion2(T,tau,err)

% ESTA FUNCION ESTIMA LAS PROBABILIDADES DE TRANSICION P(t)
% EN LA MATRIZ "Pt", EN EL INTERVALO DE TIEMPO [0,tau] CON
% ERROR DE TOLERANCIA "err". UNA VARIANTE DEL METODO ES DE
% SE PUEDE CALCULAR LA PRIMERA TRANSICION CON UN ERROR "err"
% BAJO Y SE UTILIZA LA PROPIEDAD DE CHAPMAN-KOLMOGOROV
%
%           P(a+b) = P(a)P(b)
%
% FORMA DE USARSE Pt = uniformizacion2(T,tau,err)

% Generamos "la malla" de la probabilidad P(t) para cada 0<t1<t2<...<tau
summat = zeros(n);
Pt = zeros(n,n,m);

% Calculamos P(t1) como antes
Poisson = poisspdf(xapo,lam*xt(2));
for q = 1:Nopt+1
    summat = summat + Poisson(q)*Pprob(:, :, q);
end
Pt(:, :, 1) = summat;

% Vamos guardando las demas probabilidades P(tk)
for k=2:m
    Pt(:, :, k) = Pt(:, :, 1)*Pt(:, :, k-1);
end
```

Quemos ver cuál de los dos métodos es más eficiente computacionalmente, debido a que en teoría nos tienen que regresar las mismas probabilidades. El siguiente código grafica fácilmente las dos probabilidades y compara por el momento de manera visual ambas probabilidades.

```
function [tiempo1,tiempo2]=ejemplo_uniformizacion(T,tau,err1,err2)

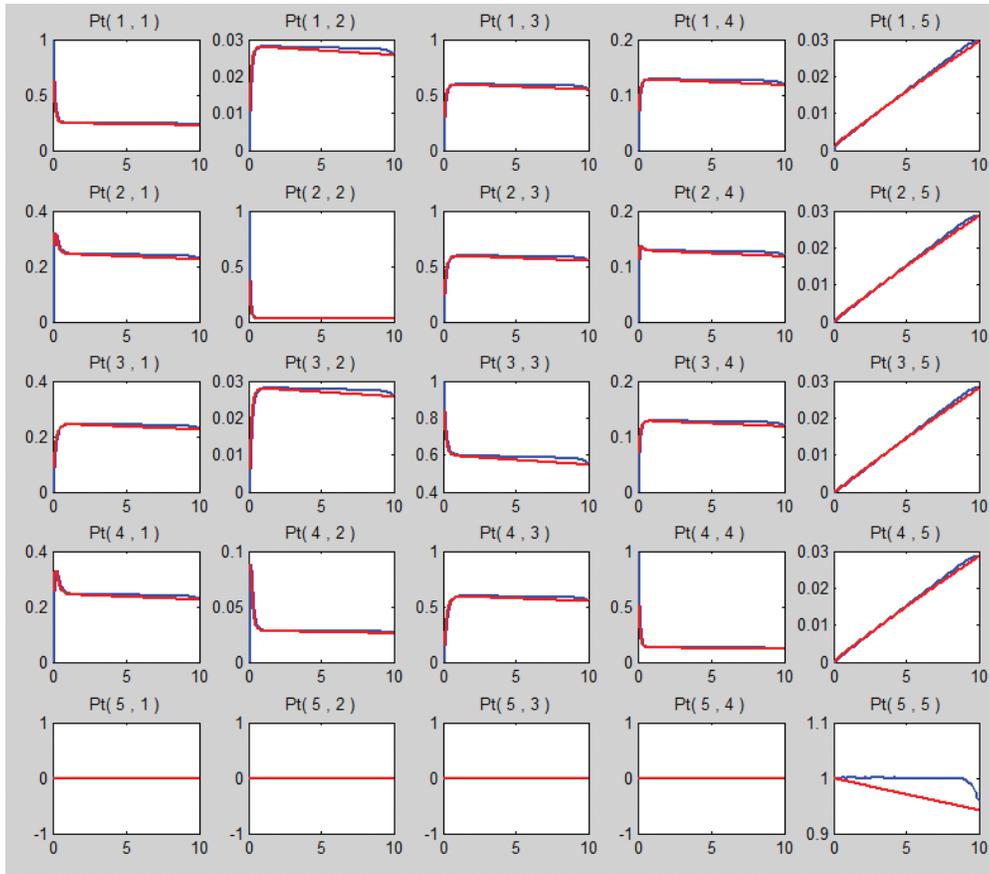
% ESTA FUNCION GRAFICA SIMULTANEAMENTE LAS PROBABILIDADES DE TRANSICION
% Pij(t) PARA CADA TIEMPO EN EL INTERVALO [0,tau], PARA PODER COMPARAR
```

```
% LAS PROBABILIDADES DE TRANSICION UTILIZANDO LAS DOS VARIANTES DEL  
% METODO DE UNIFORMIZACION CON ERRORES "err1" y "err2".
```

```
tic, Pt1 = uniformizacion(T,tau,err1); tiempo1=toc;  
tic, Pt2 = uniformizacion2(T,tau,err2); tiempo2=toc;  
  
xt = 0:0.1:tau;  
n=max(size(T));  
  
cont=1;  
  
for i = 1:n  
    for j = 1:n  
        vecPij1=copia_valores_multidimensional(Pt1,i,j);  
        vecPij2=copia_valores_multidimensional(Pt2,i,j);  
        subplot(n,n,cont), plot(xt,vecPij1,'b',xt,vecPij2,'r')  
        title(sprintf(' Pt( %d , %d ) ', i,j ))  
        cont=cont+1;  
    end  
end
```

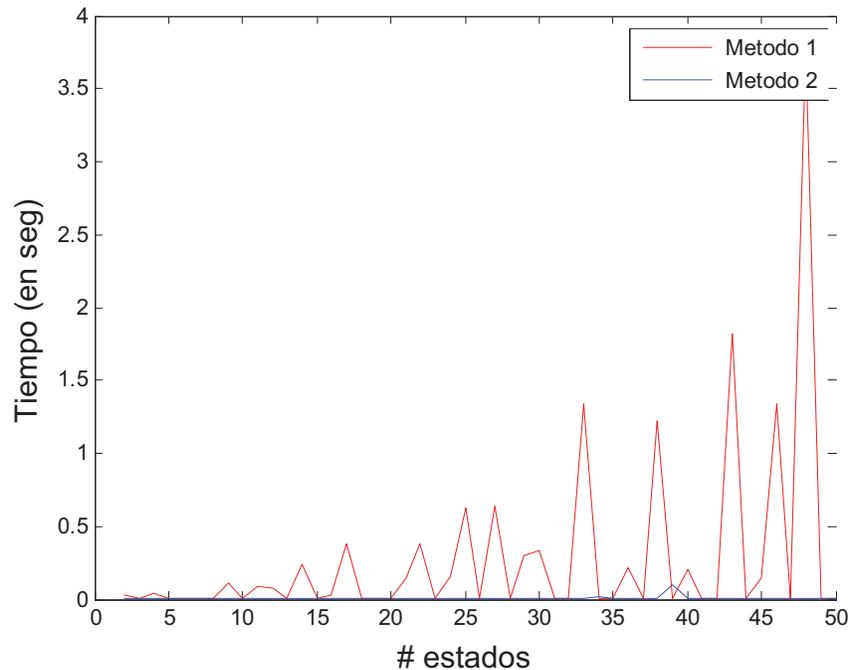
Comparando ambos métodos

```
>> [tiempo1,tiempo2]=ejemplo_uniformizacion(T,10,0.05,0.001)  
tiempo1 =  
    0.0564  
  
tiempo2 =  
    0.0013
```



Si observamos, las probabilidades son muy parecidas, salvo el caso raro de la probabilidad  $P_{5,5}(t)$  ya que el estado 5 es un estado recurrente absorbente y los demás estados son transitorios, por lo que se debería cumplir que  $P_{5,5}(t) = 1 \forall t \in \mathbb{R}^+$ . Entonces, de manera particular, podemos observar que el método 2 es bueno para estados que no son absorbentes, y además tiene la *ventaja* de que el tiempo de cálculo es relativamente menor en el método 2 que con el método 1. Esto debido a que solamente hay que calcular una sola gran malla de matrices para calcular  $P_{i,j}(0.1)$  y después son solamente multiplicaciones de matrices las que hay que realizar, algo que Mat/Oct en sus rutinas internas trae optimizado. La desventaja del método 1 es que hay que estar calculando diferentes mallas para cada  $t^* \in [0, \tau]$  por lo que requiere más tiempo memoria para estar almacenando todas estas matrices. Aquí se muestra una gráfica básica con los tiempos de cálculo para matrices de tasas de intensidad  $\mathbf{T} \in \mathbb{R}^{n \times n}$ ,  $n = 2, 3, \dots, 50$  aunque en este caso todas las matrices de tasas representan a procesos irreducibles, donde solo hay una sola clase de comunicación y no existen distintas

clases de comunicación ni estados transitorios.



Podemos observar que el método 2 es numéricamente más eficiente debido a que el tiempo de cálculo es mucho menor que con el método 1 de uniformización. La pregunta que debemos hacer es ¿cuál método nos regresa mejores resultados? En ambos casos podemos disminuir el error de tolerancia  $\varepsilon^*$  y obviamente obtendremos mejores resultados, pero sabemos que eso puede aumentar considerablemente el tiempo de cálculo, por lo que hay que tomar una decisión de que método utilizar. Para hacer una comparación un poco más exacta, utilizaremos matrices de tasas donde los valores propios son todos distintos para que la matriz pueda ser diagonalizable, y utilizar la propiedad

$$\mathbf{P}(t) = \mathbf{V}e^{\mathbf{D}t}\mathbf{V}^{-1}.$$

Con el ejemplo anterior, se tiene que:

```
>> T
T =
-5.1100    0.1000    4.0000    1.0000    0.0100
 5.0000  -10.0000    3.0000    2.0000    0
 1.0000     0    -2.0010    1.0000    0.0010
 4.0000    2.0000    1.0000   -7.0000    0
 0         0         0         0         0
```

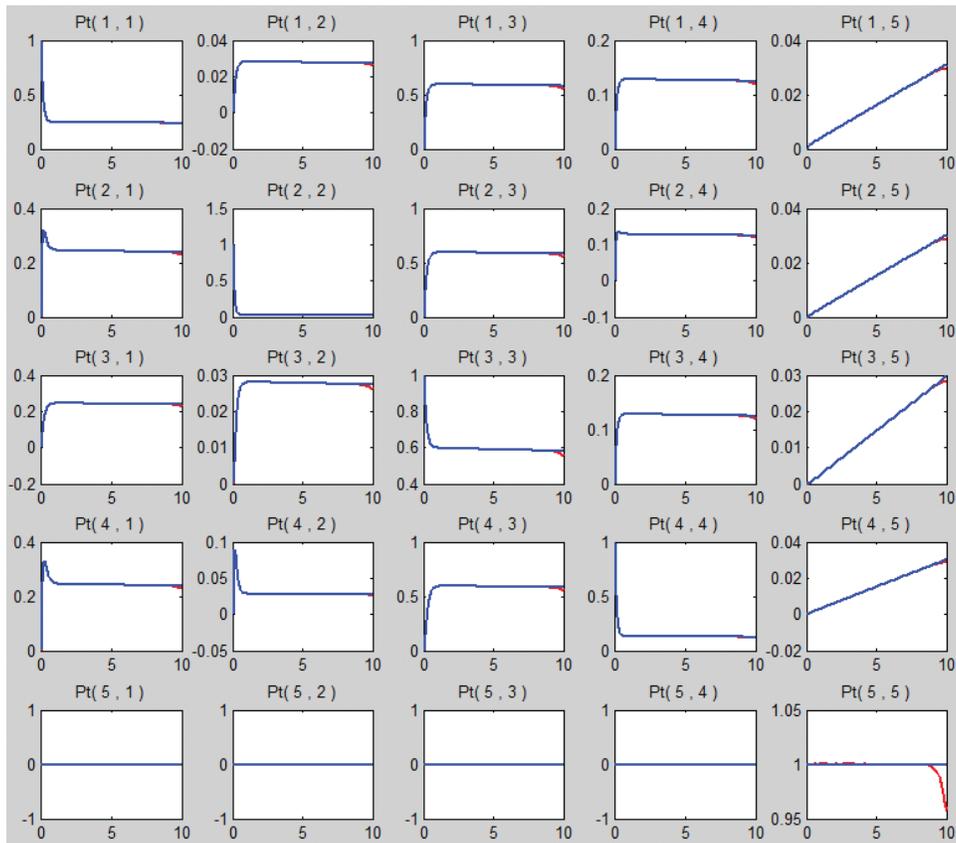
```
>> [V,D]=eig(T)
V =
-0.4996   -0.1089    0.1733   -0.0365    0.4472
-0.5001   -0.3813   -0.2296   -0.8593    0.4472
```

-0.5003	0.2525	0.1433	-0.0536	0.4472
-0.5000	-0.8826	-0.9469	0.5073	0.4472
0	0	0	0	0.4472

D =

-0.0031	0	0	0	0
0	-5.9284	0	0	0
0	0	-7.3984	0	0
0	0	0	-10.7812	0
0	0	0	0	0

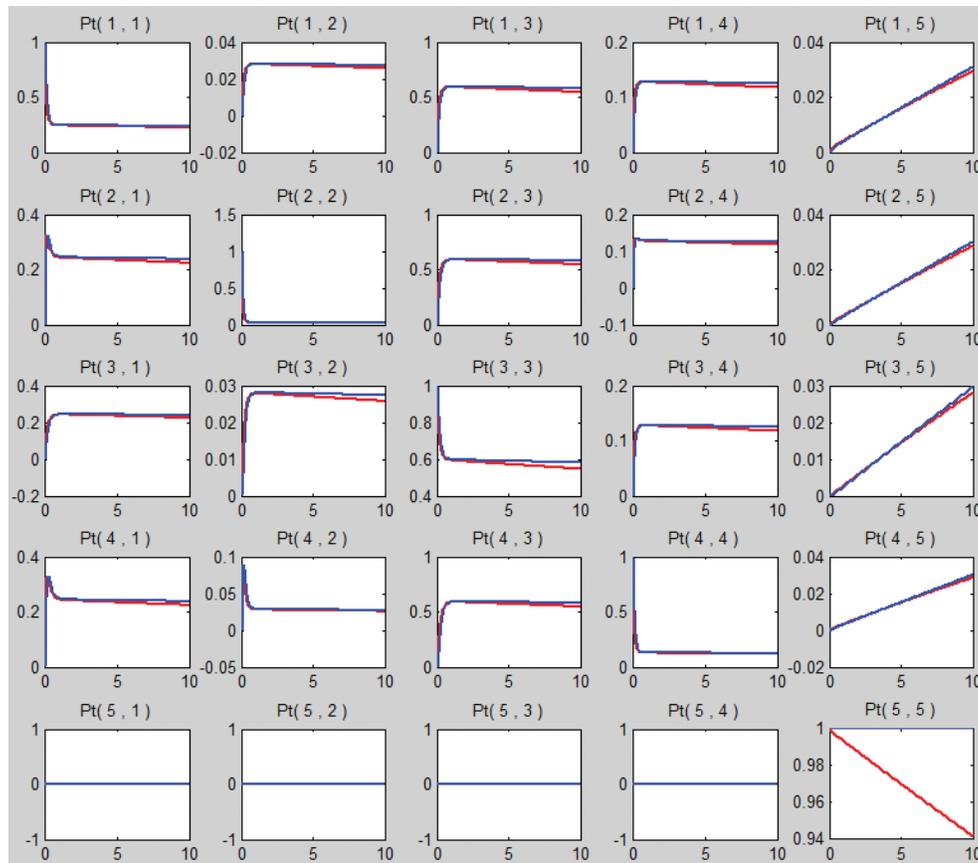
Los valores propios de la matriz son todos distintos, y esto nos permite usar el método de diagonalización. A continuación se muestran gráficamente la comparación visual de los resultados obtenidos (en rojo se muestran los resultados con el método de uniformización y en azul los resultados del método de diagonalización).



Uniformizacion vs Diagonalizacion  $\tau = 10$  y  $\varepsilon = 0.05$

No se alcanza a apreciar una diferencia significativa en los resultados de ambos métodos porque nos están regresando valores casi iguales, salvo en la última entrada  $P_{5,5}(t)$ , donde resulta mejor calcular las probabilidades de estados absorbentes con el método de diagonalización. Estos resultados nos dicen que el método de uniformización nos da buenos resultados,

aunque ya se había observado que habrá problemas con los estados que sean absorbentes, por lo que estos estados habrá que poder descartarlos desde un principio. Ahora se muestran los resultados usando la modificación del método de uniformización.



Uniformización 2 vs Diagonalización  $\tau = 10$  y  $\varepsilon = 0.001$

Se alcanza a apreciar de manera visual que la modificación, aunque es buena por la rapidez que gana en su ejecución, no es tan buena por la diferencia que hay entre la probabilidad estimada por el segundo método de uniformización y el método de diagonalización es significativa, en algunas de las entradas, en este caso particular  $P_{1,3}(t)$ ,  $P_{3,3}(t)$  y  $P_{5,5}(t)$ . En este caso se debe disminuir aún más el error permitido para poder encontrar probabilidades aceptables, y en caso de saber de antemano que tenemos estados absorbentes pues deberemos "ignorar" por un momento, el error cometido al estimar sus probabilidades de transición.

## 2.4 Método de Padé

Una de las formas de calcular las probabilidades de transición está dada por la función

$$\mathbf{P}(t) = e^{\mathbf{A}t} \quad \text{con } t \in \mathbb{R}^+$$

Recordando la expresión (2.6) la matriz exponencial se define como la serie de potencias matricial

$$e^{\mathbf{A}} = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \dots \text{ con } \mathbf{A} \in \mathbb{R}^{n \times n}$$

Con esta definición se trabajaron previamente algunas propiedades básicas de la matriz exponencial. Existe en la literatura una función racional que es una aproximación a la función exponencial conocida como **aproximación de Padé**, que fue desarrollada por el matemático Henri Padé<sup>5</sup>. Esta aproximación racional consiste de dos valores enteros positivos,  $p$  y  $q$ , que en realidad son los exponentes máximos del numerador y denominador de la siguiente función

$$R_{pq}(x) = \frac{N_{pq}(x)}{D_{pq}(x)} \text{ con } x \in \mathbb{R} \tag{2.26}$$

En Burden [18] se explica que una de las ventajas de utilizar el método de Padé es de que si se tiene grados  $p$  y  $q$  similares la aproximación es mucho mejor que si se toma una aproximación polinomial con un costo computacional similar. Esta función,  $R_{pq}(x)$ , coincide con la expansión en series de Taylor de la función exponencial hasta el término  $p+q$ , es decir, se trunca la expansión hasta ese término

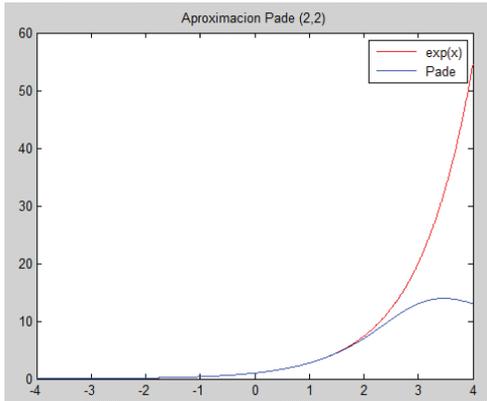
$$\sum_{k=1}^{\infty} \frac{x^k}{k!} - R_{pq}(x) = O(x^{p+q+1})$$

En Stewart [13] se muestra cuál es la forma explícita que tienen tanto el numerador como el denominador cuando se quiere encontrar la forma de Padé ( $p, q$ ) de la función exponencial  $f(\mathbf{X}) = e^{\mathbf{X}}$ . Estas formas explícitas son

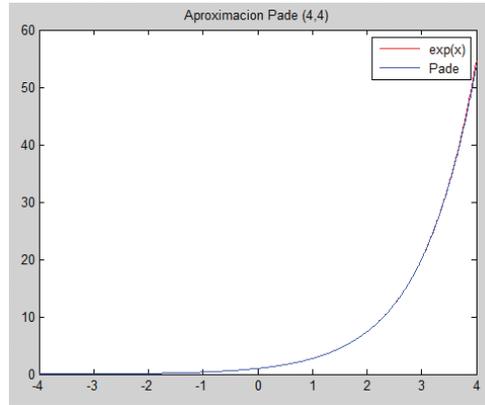
$$\begin{aligned} N_{pq}(\mathbf{A}) &= \sum_{k=0}^p \frac{(p+q-k)!p!}{(p+q)!k!(p-k)!} \mathbf{A}^k \\ D_{pq}(\mathbf{A}) &= \sum_{k=0}^q \frac{(p+q-k)!q!}{(p+q)!k!(q-k)!} (-\mathbf{A})^k \end{aligned} \tag{2.27}$$

---

<sup>5</sup>Henri Padé (1863-1953). Matemático francés. Una de sus contribuciones más importantes fue el desarrollo de las técnicas de aproximación usando funciones racionales. Su trabajo doctoral se basó en lo que la actualidad conocemos como "aproximación de Padé".



$$R_{2,2}(x) = \frac{\frac{1}{12}x^2 + \frac{1}{2}x + 1}{\frac{1}{12}x^2 - \frac{1}{2}x + 1}$$



$$R_{4,4}(x) = \frac{\frac{1}{1680}x^4 + \frac{1}{84}x^3 + \frac{3}{28}x^2 + \frac{1}{2}x + 1}{\frac{1}{1680}x^4 - \frac{1}{84}x^3 + \frac{3}{28}x^2 - \frac{1}{2}x + 1}$$

También en Stewart [13] se recomienda utilizar la aproximación diagonal de Padé, esto solamente implica utilizar a  $p = q$  y además se explota una relación que hay entre  $N_{pq}(\mathbf{A})$  y  $D_{pq}(\mathbf{A})$ . Se mostrarán a continuación de manera explícita los resultados básicos que están descritos en [13] página 420

**Elementos básicos para calcular la matriz exponencial  $e^{\mathbf{A}}$**

1)  $R_{pp}(\mathbf{A}) = N_{pp}^{-1}(-\mathbf{A})N_{pp}(\mathbf{A})$

2)  $N_{pp}(\mathbf{A}) = \sum_{j=0}^p \frac{(2p-j)!p!}{(2p)!j!(p-j)!} \mathbf{A}^j = \sum_{j=0}^p c_j \mathbf{A}^j$  (2.28)

3)  $c_0 = 1$  y  $c_j = c_{j-1} \left[ \frac{p+1-j}{2p+1-j} \right]$  con  $j = 1, 2, \dots, p$ .

El método efectivo para calcular estas aproximaciones (2.28) se muestra a continuación, aquí

hay que diferenciar si el grado de la aproximación de Padé es par o impar

$$R_{pp}(\mathbf{A}) = \begin{cases} \mathbf{I} + 2 \frac{\mathbf{A} \left[ \sum_{k=0}^{p/2-1} c_{2k+1} \mathbf{A}^{2k} \right]}{\sum_{k=0}^{p/2} c_{2k} \mathbf{A}^{2k} - \mathbf{A} \left[ \sum_{k=0}^{p/2-1} c_{2k+1} \mathbf{A}^{2k} \right]} & \text{si } p \text{ es par} \\ -\mathbf{I} - 2 \frac{\sum_{k=0}^{(p-1)/2} c_{2k} \mathbf{A}^{2k}}{\mathbf{A} \left[ \sum_{k=0}^{(p-1)/2} c_{2k+1} \mathbf{A}^{2k} \right] - \sum_{k=0}^{(p-1)/2} c_{2k} \mathbf{A}^{2k}} & \text{si } p \text{ es impar.} \end{cases}$$

El autor explica que con este método se obtienen mayores ahorros computacionales en tiempo debido a que solamente se necesitan hacer la mitad de las operaciones que si se evaluara directamente la expresión (2.28) pero que tiene la desventaja que utiliza más memoria debido a que hay que almacenar algunas matrices extra. Todo esto se reduce a la siguiente manera más compacta de escribirlo

$$\begin{aligned} 1) \text{ Si } p \text{ es par entonces } R_{pp}(\mathbf{A}) &= \mathbf{I} + 2(\mathbf{T}_{\text{par}} - \mathbf{S}_{\text{par}})^{-1} \mathbf{S}_{\text{par}} \text{ con} \\ &\quad \mathbf{S}_{\text{par}} = c_1 \mathbf{A} + c_3 \mathbf{A}^3 + \dots + c_{p-1} \mathbf{A}^{p-1} \\ &\quad \mathbf{T}_{\text{par}} = c_0 \mathbf{I} + c_2 \mathbf{A}^2 + c_4 \mathbf{A}^4 + \dots + c_p \mathbf{A}^p \\ 2) \text{ Si } p \text{ es impar entonces } R_{pp}(\mathbf{A}) &= -\mathbf{I} - 2(\mathbf{T}_{\text{impar}} - \mathbf{S}_{\text{impar}})^{-1} \mathbf{S}_{\text{impar}} \text{ con} \\ &\quad \mathbf{S}_{\text{impar}} = c_0 \mathbf{I} + c_2 \mathbf{A}^2 + c_4 \mathbf{A}^4 + \dots + c_{p-1} \mathbf{A}^{p-1} \\ &\quad \mathbf{T}_{\text{impar}} = c_1 \mathbf{A} + c_3 \mathbf{A}^3 + \dots + c_p \mathbf{A}^p. \end{aligned}$$

Lo que hace falta comentar es que este procedimiento del método de Padé hace un uso explícito de que la norma de la matriz  $\mathbf{A}$  es "pequeña", lo que significa en teoría es que solamente se pueden trabajar con matrices  $\mathbf{A} \approx \mathbf{0}$ , pero esto ocurre rara vez en la práctica. Para ello entonces se utiliza la siguiente relación matricial

$$e^{\mathbf{A}} = (e^{\mathbf{A}/m})^m \tag{2.29}$$

y de preferencia se utiliza una potencia de 2, es decir,  $m = 2^s$ , esto con la finalidad de que al realizar las multiplicaciones matriciales uno pueda ahorrar trabajo computacional. Por eso a este método en inglés se le conoce como **scaling and squaring** y que en español

una traducción literal podría ser **escalamiento y potencia** (primero se realiza la escala para tener una normal matricial pequeña y después se regresa a la matriz original usando potencias sucesivas). Esta técnica del escalamiento y potencia es muy útil porque permite calcular potencias de matrices grandes sin tener que realizar muchas multiplicaciones de matrices

$$\begin{aligned} \mathbf{A}^2 &= \mathbf{A}\mathbf{A} \\ \mathbf{A}^4 &= \mathbf{A}^2\mathbf{A}^2 \\ \mathbf{A}^8 &= \mathbf{A}^4\mathbf{A}^4 \\ &\vdots \end{aligned}$$

En Stewart [13] el autor dice que el valor  $m$  óptimo está dado por

$$m_{opt} = \max \left\{ 0, \left\lfloor \frac{\log \|\mathbf{A}\|_{\infty}}{\log 2} \right\rfloor + 1 \right\}. \quad (2.30)$$

A continuación se muestra la implementación del código propuesto por el autor para encontrar la exponencial de la matriz.

```
function R=exponencial_pade(X,p)

% ESTA FUNCION CALCULA UNA APROXIMACION A LA MATRIZ EXPONENCIAL
% DE LA MATRIZ "X" USANDO LA APROXIMACION DE PADE CON VALOR
% "p" DE VALORES. EL ALGORITMO ES EL QUE APARECE EN EL LIBRO
% "INTRODUCTION TO THE NUMERICAL SOLUTION OF MARKOV CHAINS" DEL AUTOR
% WILLIAM STEWART, PAG 421
%
% FORMA DE USARSE: R = exponencial_pade(X,p)

% Detectar cual es el valor optimo
mopt = max(0,fix(log(norm(X,'inf'))/log(2)) + 1);

% Se calculan los coeficientes de la aproximacion de Pade
c(1) = 1;
for k = 1:p
    c(k+1) = c(k)*((p+1-k)/(k*(2*p+1-k)));
end

% Se hace el escalamiento para tener una norma matricial pequena
% y las matrices auxiliares para el calculo
X1=power(2,-mopt)*X; X2=X1*X1; I=eye(size(X));
T=c(p+1)*I; S=c(p)*I; impar=1;
```

```

% La evaluacion polinomial matricial
for k = p-1:-1:1
    if( impar == 1 )
        T = T*X2 + c(k)*I;
    else
        S = S*X2 + c(k)*I;
    end
    impar=1-impar;
end

if( impar==0 )
    S=S*X1;
    R=I + 2*((T-S)\S);
else
    T=T*X1;
    R=-(I + 2*((T-S)\S));
end

% Sacamos potencias sucesivas
for j=1:mopt
    R=R*R;
end

```

Una de las conclusiones iniciales en Stewart [13] con respecto a que valor de  $p$  tomar para el método de Padé es tomar un valor de  $p \approx 6$ . Esto se debe a que si el valor  $m_{opt}$  para usar el método de escalamiento y potencia es mayor que 1, entonces las potencias sucesivas se pueden ver contaminadas por los errores de redondeo y el error crecerá considerablemente. Con este archivo central se construirá otro que nos permita encontrar las probabilidades de transición usando este método y poder compararlo con los métodos de diagonalización, uniformización y el de Padé. También en este caso para las probabilidades de transición se usarán las ecuaciones de Chapman-Komogorov  $\mathbf{P}(t_1 + t_2) = \mathbf{P}(t_1)\mathbf{P}(t_2)$ , con lo que solamente será necesario calcular la matriz  $\mathbf{P}(t_1)$  con un grado alto de precisión  $p$  y después realizar las multiplicaciones sucesivas.

```

function Pt = metodo_Pade(T,p,tau)

% ESTA FUNCION CALCULA LAS PROBABILIDADES DE TRANSICION DEL
% PROCESO ESTOCASTICO CON MATRIZ DE TASAS "T" HASTA EL TIEMPO
% "tau", USANDO EL METODO DE PADE DE GRADO "p"
%
% FORMA DE USARSE: Pt = metodo_Pade(T,p,tau)

[n1,n2]=size(T);
if(n1~=n2)
    error(' LA MATRIZ DEBE SER CUADRADA!!! ')
end

```

```

end

n=n1;

% Nos aseguramos que el valor de "p" sea entero y positivo
p=abs(round(p));

% Construimos la particion del tiempo
xt = 0:0.1:tau; m = length(xt);

% Calculamos las probabilidades de transicion P(t) con Pade
Pt = zeros(n,n,m); Pt(:,:,1)=exponencial_pade(0.1*T,p);

for k=2:m
    Pt(:,:,k) = Pt(:,:,k-1)*Pt(:,:,1);
end

% Por ultimo, graficamos estas probabilidades para cada (i,j)
cont=1;
for i = 1:n
    for j = 1:n
        vecPij=copia_valores_multidimensional(Pt,i,j);
        subplot(n,n,cont), plot(xt,vecPij)
        title(sprintf(' Pt( %d , %d ) ', i,j ))
        cont=cont+1;
    end
end
end

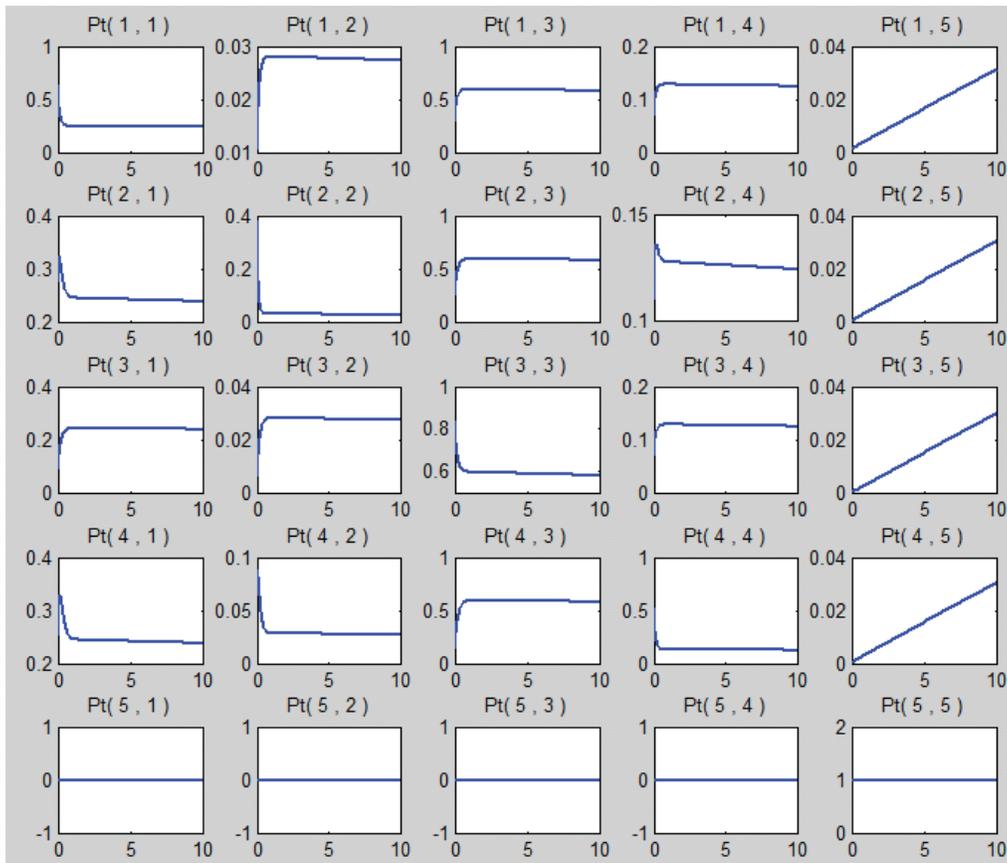
```

Utilizando este método para la matriz de tasas de intensidad (2.25) obtenemos los siguientes resultados.

```

>> T=[-5.11 0.1 4 1 0.01; 5 -10 3 2 0; 1 0 -2.001 1 0.001; 4 2 1 -7 0; 0 0 0 0 0];
>> Pt = metodo_Pade(T,6,10);

```



Se observa que este método es muy bueno debido a que las probabilidades de transición del estado 5 absorbente son 1, tal como se espera. Otra cosa a destacar es que con este método se evita calcular los valores y vectores propios de una matriz, y el tiempo de cómputo es relativamente menor que con los métodos vistos previamente. Veremos cuál de los métodos es el más eficiente en tiempo computacional, en este sentido solamente es para tener una referencia sobre el tiempo que se tarda en calcular el CPU las matrices de transición. Para ello se construirá una matriz "aleatoria" de tasas de transición  $\mathbf{T} \in \mathbb{R}^{n \times n}$  para comparar los métodos antes vistos. La idea central es construir una matriz de tasas de un proceso irreducible, donde solo haya una sola clase de comunicación, las tasas serán aleatorias para  $i \rightarrow j$ , y al final sumaremos estos valores y los pondremos en la diagonal principal.

**Algoritmo para generar una matriz de tasas**

- 1) Seleccionar el número de estados  $n$  y la tasa  $\lambda$
- 2) Para  $i \neq j$  generar  $T_{i,j} \sim U(0, \lambda)$
- 3) Generar  $T_{i,i} = -\sum_{j \neq i} T_{i,j}$

```
function T=contruccion_matriz_tasas(n,lambd)
```

```
% ESTA FUNCION CONSTRUYE UNA MATRIZ DE TASAS DE
% TRANSICION DE UN PROCESO CON "N" ESTADOS Y UNA
% TASA GENERADORA ALEATORIA "lambda"
%
% FORMA DE USARSE: T=contruccion_matriz_tasas(n,lambda)
```

```
% Construimos la matriz de tasas a partir de v.a.
% uniformes  $T(i,j) \sim U(0,L)$ 
T=unifrnd(0,lambda,n,n);
```

```
% La diagonal la convertimos en 0's
for k=1:n
    T(k,k)=0;
end
```

```
%Calculamos la suma por renglones
t=sum(T,2);
```

```
% La diagonal es la suma por renglon
for k=1:n
    T(k,k)=-t(k);
end
```

Ahora se muestra el código para estimar los tiempos que se tardan ambos métodos en calcular la matriz exponencial.

```
function [tiempos2,tiempos3,tiempos4]= analisis_tiempo_matrices(N,tau,err2,p)
```

```
% Aqui almacenaremos los tiempos
tiempos2=zeros(1,N-1); tiempos3=zeros(1,N-1); tiempos4=zeros(1,N-1);
```

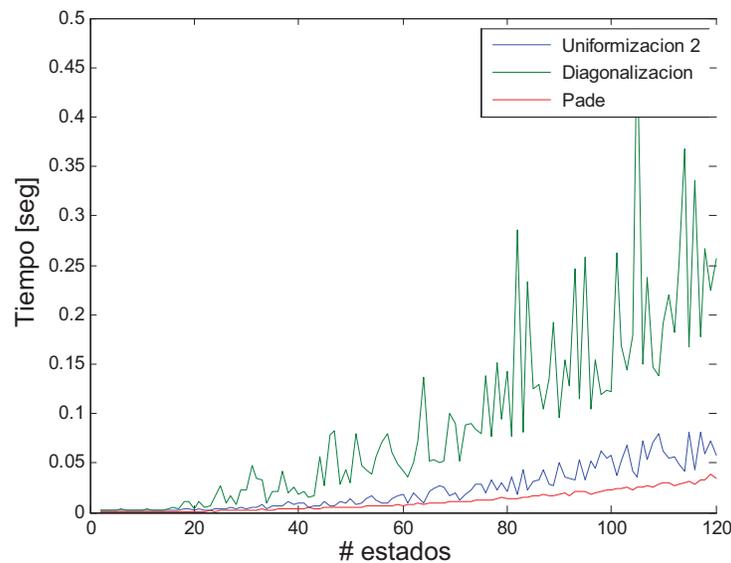
```
for k=1:N-1
    T=contruccion_matriz_tasas(k+1,unifrnd(1,20));
    [tiempo2,tiempo3,tiempo4]=tiempo_metodos(T,tau,err1,err2,p);
    tiempos2(k)=tiempo2; tiempos3(k)=tiempo3; tiempos4(k)=tiempo4;
end
```

```
plot(2:N,tiempos2,2:N,tiempos3,2:N,tiempos4)
legend(' Uniformizacion 2 ', ' Diagonalizacion ', ' Pade ')
```

```
function [tiempo2,tiempo3,tiempo4]=tiempo_metodos(T,tau,err2,p)
```

```
tic, Pt = uniformizacion2(T,tau,err2); tiempo2=toc;
tic, Pt = diagonalizacion(T,tau); tiempo3=toc;
tic, Pt = metodo_Pade(T,p,tau); tiempo4=toc;
```

Este programa m-file construye una matriz de tasas de transición "aleatorias" de un orden especificado, después calcula las matrices exponenciales con los métodos de uniformización, diagonalización y Padé, y va registrando los tiempos que se tarda el CPU en calcular estas matrices y por último regresa la gráfica de los tiempos. A continuación se muestra cuanto tiempo se tarda en Mat/Oct cada uno de los métodos.



Podemos apreciar empíricamente que los tiempos van aumentando cuando se tienen matrices de tasas cada vez mayores (como se esperaría), pero se observa que el método de diagonalización no es muy bueno, debido a que el tiempo es considerablemente mayor, esto porque el sistema inicialmente tiene que estar calculando los vectores y valores propios, y después se tienen que hacer tres multiplicaciones de matrices por iteración. El método de Padé con  $p = 6$  propuesto en Stewart [13] es realmente eficiente en tiempo computacional, debido a que para matrices de 120 estados en menos de un segundo calculó las probabilidades de transición para  $\tau \in [0, 10]$ , y los resultados numéricos son buenos, por lo que ya podremos considerar este método como base para realizar los cálculos matriciales. Por último se verá un método de Padé mejorado, analizado a fondo en Higham [8], donde se muestra una versión superior a este método de Padé revisado anteriormente.

## 2.5 Método de Padé mejorado

Se expondrá aquí el método de Padé mejorado, analizado a fondo en Higham [8], donde analizan cuestiones del error generado al calcular la función exponencial usando el método de Padé, cotas para el error y un algoritmo óptimo para encontrar la matriz exponencial usando aritmética de doble precisión. Primero el autor da expresiones para el error escalar y matricial sobre los errores generados entre el valor exacto de la exponencial y usando el

truncamiento del método de Padé

$$\begin{aligned}
 e^x - R_{pq}(x) &= (-1)^q \frac{p!q!}{(p+q)!(p+q+1)!} x^{p+q+1} + O(x^{p+q+2}) \\
 e^{\mathbf{A}} - R_{pq}(\mathbf{A}) &= \frac{(-1)^q}{(p+q)!} \mathbf{A}^{p+q+1} [D_{pq}(\mathbf{A})]^{-1} \int_0^1 e^{\mathbf{A}t} (1-t)^{p+q} dt.
 \end{aligned}
 \tag{2.31}$$

También se argumenta que las aproximaciones de Padé deben ser del mismo tamaño  $p = q$  debido a que  $R_{pp}(\cdot)$  se puede evaluar con el mismo costo computacional. El procedimiento que el autor propone es que se busca un valor  $s$  para realizar el escalamiento tal que se pueda acotar el error en términos de la norma matricial  $\|2^{-s} \mathbf{A}\|$ , y después para cada grado  $m$  encontrar el máximo  $\|2^{-s} \mathbf{A}\|$  para que la aproximación  $R_{pp}(\cdot)$  se pueda aproximar bien el cálculo de la matriz exponencial. Se busca que

$$e^{-\mathbf{A}} R_{pp}(\mathbf{A}) = \mathbf{I} + \mathbf{G} = e^{\mathbf{H}}.$$

es decir, que la aproximación de Padé por la inversa de exponencial sea la matriz identidad más una pequeña perturbación  $\mathbf{G}$  tal que  $\|\mathbf{G}\| < 1$  y describirlo como  $\mathbf{H} = \log(\mathbf{I} + \mathbf{G})$ . Uno de los resultados clave del autor los enuncia es el siguiente teorema.

**Teorema 2.15 (Higham)** *Sea la aproximación de Padé diagonal que satisfaga la relación*

$$e^{-2^{-s} \mathbf{A}} R_{pp}(2^{-s} \mathbf{A}) = \mathbf{I} + \mathbf{G}$$

donde la matriz  $\mathbf{G}$  tiene una norma matricial  $\|\mathbf{G}\| < 1$  y la norma es cualquier norma matricial. Entonces se cumple

$$[R_{pp}(2^{-s} \mathbf{A})]^{2^s} = e^{\mathbf{A} + \mathbf{E}}$$

donde la matriz  $\mathbf{E}$  conmuta con la matriz  $\mathbf{A}$  y se cumple

$$\frac{\|\mathbf{E}\|}{\|\mathbf{A}\|} \leq \frac{-\log(1 - \|\mathbf{G}\|)}{\|2^{-s} \mathbf{A}\|} \tag{2.32}$$

En un análisis realizado por el autor, explica que para poder dar una cota para la matriz  $\mathbf{G}$  se debe analizar la función

$$\rho(x) = e^{-x} R_{pp}(x) - 1$$

que es la diferencia entre la aproximación de Padé y la exponencial. Como se debe trabajar con el truncamiento de la aproximación de Padé entonces esta función  $\rho(\cdot)$  debe cumplir también que

$$\rho(x) = \sum_{k=2p+1}^{\infty} c_k x^k$$

y que esta serie convergerá si se cumple que  $|x| < v_p = \min \{|t| \mid D_{pp}(t) = 0\}$ , y entonces se cumple que  $\|\mathbf{G}\| \leq \sum_{k=2p+1}^{\infty} |c_k| \theta^k = f(\theta)$ , donde  $\theta = \|2^{-p}\mathbf{A}\| < v_p$ . El autor analizó esta serie (2.32) para que esta no exceda la unidad de redondeo  $u = 2^{-53} \approx 1.1 \times 10^{-16}$  al trabajar en doble precisión. Esto es importante para el desarrollo de la mejora del método de Padé. También con estos datos encuentra que el valor óptimo para máquinas de doble precisión es tomar una aproximación con  $p = 13$  y obtiene que  $\theta_{13} = 5.4$ , ya que con su análisis demuestra que solamente son necesarias 6 multiplicaciones matriciales para lograr una aproximación con 13 términos. El autor demuestra que los coeficientes en el numerador en la aproximación  $P_m(x) = \sum_{k=0}^m b_k x^k$  son

$$\begin{array}{ll} b_0 = 64764752532480000 & b_7 = 33522128640 \\ b_1 = 32382376266240000 & b_8 = 1323241920 \\ b_2 = 7771770303897600 & b_9 = 40840800 \\ b_3 = 1187353796428800 & b_{10} = 960960 \\ b_4 = 129060195264000 & b_{11} = 16380 \\ b_5 = 10559470521600 & b_{12} = 182 \\ b_6 = 670442572800 & b_{13} = 1 \end{array}$$

A continuación se muestra una implementación en Mat/Oct del algoritmo propuesto en Higham [8] página 246 para encontrar la matriz exponencial de cualquier matriz  $\mathbf{A}$ , especializado para aritmética de doble precisión.

```
function X = pade_Higham(A)

% ESTA FUNCION ENCUENTRA LA MATRIZ EXPONENCIAL DE LA MATRIZ "A"
% UTILIZANDO EL ALGORITMO PROPUESTO POR NICHOLAS HIGHAM EN EL LIBRO
% "FUNCTIONS OF MATRICES. THEORY AND COMPUTATION", QUE ES UNA
% MEJORA EL METODO DE PADE.
%
% FORMA DE USARSE: X = pade_Higham(A)

% Se toman los valores especificados en el articulo de Higham
b = [64764752532480000, 32382376266240000, 7771770303897600,...
    1187353796428800, 129060195264000, 10559470521600,...
    670442572800, 33522128640, 1323241920, 40840800,...
    960960, 16380, 182, 1];

% Se verifica que sea una matriz cuadrada
```

```

[n,m]=size(A);
if( n ~= m )
    error('LA MATRIZ "A" DEBE SER CUADRADA');
end

% Se normaliza la matriz para tener norma pequena
s = ceil( log2( norm(A,1) / 5.4 ) );
A = A/(2^s);
Id = eye(n);

% Se hace la construccion de las matrices auxiliares
A2=A*A; A4=A2*A2; A6=A2*A4;
U=A*( A6*(b(14)*A6 + b(12)*A4 + b(10)*A2) + b(8)*A6 + b(6)*A4 + b(4)*A2 + b(2)*Id );
V=A6*(b(13)*A6 + b(11)*A4 + b(9)*A2) + b(7)*A6 + b(5)*A4 + b(3)*A2 + b(1)*Id;

% Se resuelve el sistema general para encontrar Rpp(A)
Rpp = (V-U)\(U+V);

% Se encuentra la matriz exponencial elevando potencias sucesivas
for k = 1:s
    Rpp = Rpp*Rpp;
end

```

Con este procedimiento básico se construye el último método para calcular la matriz exponencial.

```
function Pt = metodo_Pade_mejorado(T,tau)
```

```

% ESTA FUNCION CALCULA LAS PROBABILIDADES DE TRANSICION DEL
% PROCESO ESTOCASTICO CON MATRIZ DE TASAS "T" HASTA EL TIEMPO
% "tau", USANDO EL METODO DE HIGHAM PARA ENCONTRAR LA MATRIZ
% EXPONENCIAL
%
% FORMA DE USARSE: Pt = metodo_Pade_mejorado(T,tau)

```

```

[n1,n2]=size(T);
if(n1~=n2)
    error(' LA MATRIZ DEBE SER CUADRADA!!! ')
end

```

```

% Construimos la particion del tiempo
xt = 0:0.1:tau;
m = length(xt);

```

```

% Calculamos las probabilidades de transicion P(t) con Pade mejorado
Pt = zeros(n1,n1,m);

```

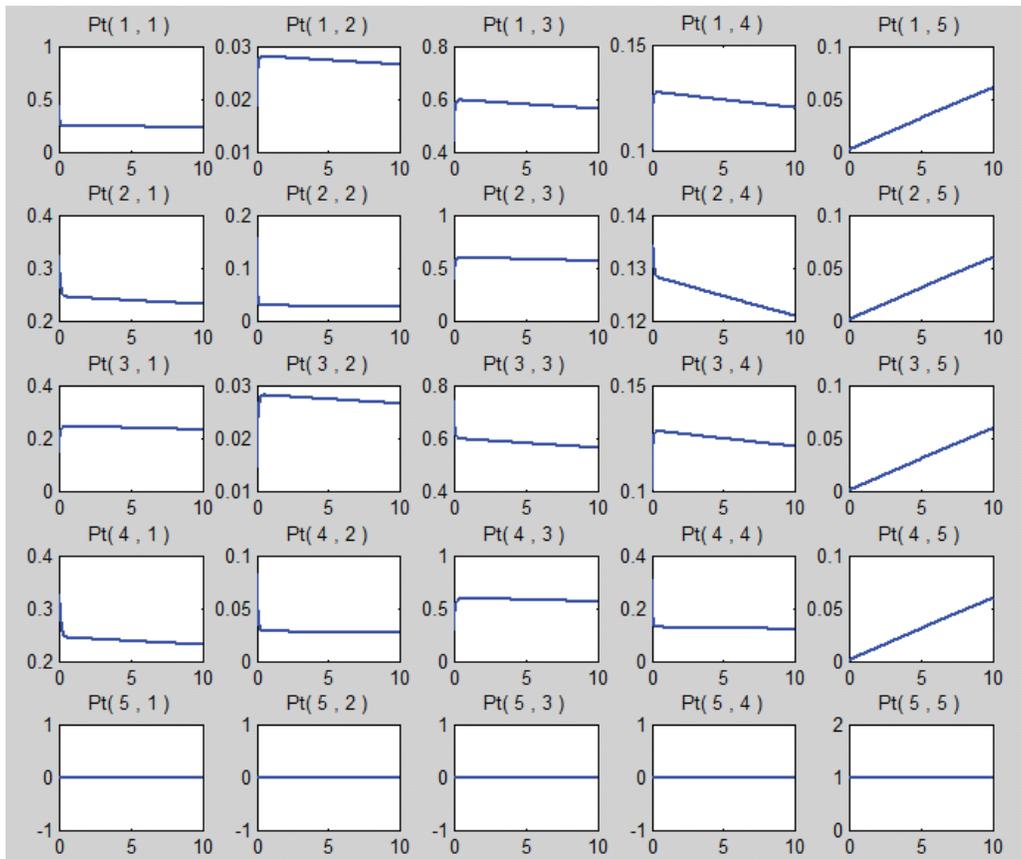
```
Pt(:,:,1)=pade_Higham(0.1*T);

for k=2:m
    Pt(:,:,k) = Pt(:,:,k-1)*Pt(:,:,1);
end

% Por ultimo, graficamos estas probabilidades para cada (i,j)
cont=1;
for i = 1:n1
    for j = 1:n1
        vecPij=copia_valores_multidimensional(Pt,i,j);
        subplot(n1,n1,cont), plot(xt,vecPij)
        title(sprintf(' Pt( %d , %d ) ', i,j ))
        cont=cont+1;
    end
end
```

Utilizando de nueva cuenta el método para la matriz de tasas de intensidad (2.25) obtenemos los siguientes resultados.

```
>> T=[-5.11 0.1 4 1 0.01; 5 -10 3 2 0; 1 0 -2.001 1 0.001; 4 2 1 -7 0; 0 0 0 0 0];
>> Pt = metodo_Pade_mejorado(T,10);
```



Este método regresa los mismos resultados que con los métodos anteriores, la ventaja del método de Padé es que no se pide ninguna estructura especial para la matriz, por lo que se puede tener una matriz de tasas de intensidad con varias clases de comunicación, con estados recurrentes y transitorios, etc. Según explica Higham [8], este método es óptimo al trabajar con máquinas que tengan una aritmética de punto flotante de doble precisión. Solamente nos faltaría comparar el tiempo en ejecución del método original de Padé con este método.

```
function [tiempo1,tiempo2]= analisis_tiempo_Pade(N,tau,p)
```

```
% Aqui almacenaremos los tiempos
```

```
tiempo1=zeros(1,N-1); tiempo2=zeros(1,N-1);
```

```
for k=1:N-1
```

```
    T=contruccion_matriz_tasas(k+1,unifrnd(1,20));
```

```
    tic, Pt=metodo_Pade(T,p,tau); tiempo1(k)=toc;
```

```
    tic, Pt=metodo_Pade_mejorado(T,tau); tiempo2(k)=toc;
```

```
end
```

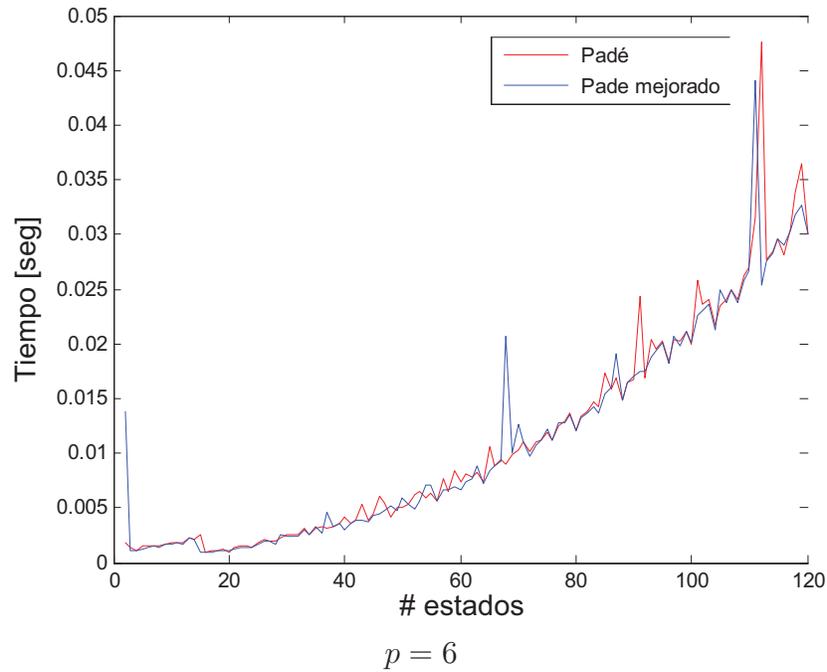
```
plot(2:N,tiempo1,'r',2:N,tiempo2,'b')
```

```
legend(' Pade ', ' Pade mejorado ')
```

Se muestran dos resultados, uno con  $p = 6$  que es el método inicial propuesto, y después

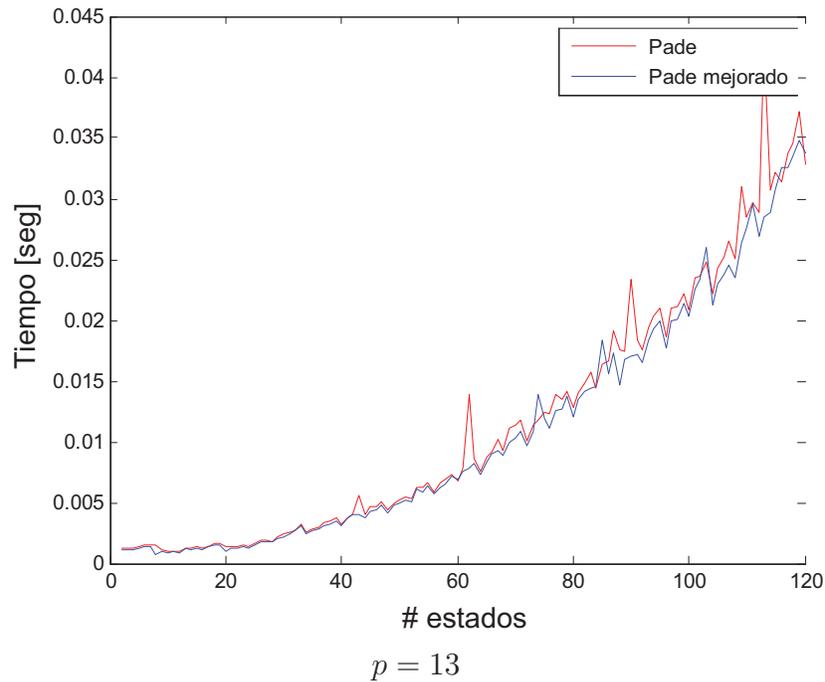
con  $p = 13$  que es el mismo grado del método de Padé mejorado.

```
>> [tiempo1,tiempo2]= analisis_tiempo_Pade(120,10,6);
```



Tenemos que el tiempo requerido es computacionalmente equivalente, pero sorprende que con el método mejorado se obtiene el mismo tiempo al encontrar  $R_{13,13}(\mathbf{A})$  que con el método

original  $R_{6,6}(\mathbf{A})$ . Se compararán ambos métodos con  $p = 13$ .



Concluyendo esta parte, al hacer cálculos en Mat/Oct se observa que este es el método que trae programado estos paquetes de cómputo para calcular la matriz exponencial, debido a que los residuales de la función `expm.m` de los paquetes mencionados anteriormente y el código propuesto es cero en todos los casos, y esto da certeza al querer encontrar en nuestro las probabilidades de transición de un proceso de Markov, y será el método que se usará en lo que resta del trabajo.

# Capítulo 3

## Teoría básica de procesos de Markov

### 3.1 Cadenas de Markov finitas

A continuación se revisarán algunos resultados importantes sobre una clase en particular de procesos estocásticos, los que son conocidos en la literatura como **procesos de Markov**.

**Definición 3.1** Sea  $(\Omega, \mathcal{A}, P)$  un espacio de probabilidad, y sea  $T \subset \mathbb{R}$  un conjunto de cardinalidad infinita. Supongamos que para cada  $t \in T$  existe una variable aleatoria  $X_t : \Omega \rightarrow \mathbb{R}$  definida en  $(\Omega, \mathcal{A}, P)$ . Entonces la función  $X : T \times \Omega \rightarrow \mathbb{R}$  denotada por  $X(t, \omega) = X_t(\omega)$  se conoce como un **proceso estocástico** con índice  $T$  y se escribe como  $\{X_t\}_{t \in T}$

El conjunto  $T$  denota por lo general al tiempo, por lo que un proceso estocástico nos dirá como fue evolucionando un sistema a través del tiempo. De la definición anterior también es importante subrayar que  $\omega$  se refiere a la realización de un evento de la variable aleatoria, por lo que si se fija un  $\omega$  quiere decir que se obtuvo una realización de todas las variables aleatorias  $\{X_t\}_{t \in T}$  y se obtiene la gráfica de una "trayectoria" del proceso. Si se fija un tiempo  $t_0$  entonces se obtendrían las posibles realizaciones de una sola variable aleatoria y se determinarían sus características. En particular una clase importante dentro de los procesos estocásticos son los procesos que no dependen de todo su pasado, a los cuales se les conoce como procesos de Markov.

**Definición 3.2** Un **proceso de Markov** es un proceso estocástico  $\{X_n\}_{n \in T}$  tal que dado el valor de  $X_n$ , el valor de  $X_m$  con  $m > n$ , no depende de los valores de  $X_k$ , con  $k < n$ .

Una nota importante es que solamente se trabajará con procesos de Markov con una cantidad finita de resultados. A este conjunto de resultados del proceso de Markov se le denota con la letra  $\mathbb{E}$  y es el **espacio de estados**, que en los casos más comunes representa un conjunto finito de números enteros  $\mathbb{E} = \{I_1, I_2, \dots, I_m\}$ , también puede representar una respuesta  $\mathbb{E} = \{\text{si}, \text{no}\}$ , entre otros. Dependiendo del contexto que se este analizando se definirá un espacio muestral adecuado. En nuestro caso solamente trabajaremos con un espacio de estados finito, para que pueda ser más práctico el análisis de algún tipo de problema en particular. Una definición más formal sobre lo que es una cadena de Markov se enuncia a continuación.

**Definición 3.3** Un proceso estocástico  $\{X_k\}_{k \in \mathbb{N}}$  definido sobre  $\Omega$  y con espacio de estados  $\mathbb{E}$  a lo más numerable es llamado **cadena de Markov a tiempo discreto** si se cumple con la propiedad de Markov

$$P[X_n = i_n \mid X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_0 = i_0] = P[X_n = i_n \mid X_{n-1} = i_{n-1}] \quad (3.1)$$

con la condición de que  $P[X_{n-1} = i_{n-1}, \dots, X_0 = i_0] > 0, \forall n \in \mathbb{N}$  y  $\forall i_0, i_1, \dots, i_n \in \mathbb{E}$ . A las probabilidades  $P[X_n = i_n \mid X_{n-1} = i_{n-1}]$  las denotaremos por  $P_{i_{n-1}, i_n}^{(n-1, n)}$ .

Esta es la propiedad que define a una cadena de Markov, el proceso se "olvida" de todo el pasado y solamente importa el pasado inmediato anterior. Esta definición recuerda en que momento se da el salto, no es lo mismo tener la probabilidad  $P[X_2 = 3 \mid X_2 = 1]$  que  $P[X_7 = 3 \mid X_6 = 1]$ , por lo que también es necesario tener una definición de procesos que no recuerden el momento y otros procesos que si lo recuerden.

**Definición 3.4** Una cadena de Markov a tiempo discreto es llamada **homogénea ó estacionaria** en el tiempo si la probabilidad de ir de un estado a otro es independiente del tiempo en el cual el paso es realizado. Esto es, para todos los estados  $i, j \in \mathbb{E}$

$$P_{i,j} = P[X_n = j \mid X_{n-1} = i] = P[X_{n+k} = j \mid X_{n+k-1} = i] \quad (3.2)$$

para  $k = -(n-1), -(n-2), \dots, -1, 0, 1, 2, \dots$ . Si la cadena de Markov no es estacionaria, entonces será llamada **no homogénea**. Si la cadena de Markov es estacionaria, para  $m \geq 1$  denotaremos por  $P_{i,j}^{(m)}$  como

$$P_{i,j}^{(m)} = P[X_{n+m} = j \mid X_n = i] \quad (3.3)$$

que es la probabilidad de ir en  $m$  pasos de  $i$  a  $j$ .

En la práctica los procesos son homogéneos, en Isaacson y Madsen [52] se muestran resultados sobre las cadenas de Markov finitas no homogéneas. La ventaja de los procesos de Markov homogéneos es que las probabilidades  $P_{i,j} = P[X_{k+1} = j \mid X_k = i]$  para cada pareja en el espacio de estados se pueden acomodar de manera matricial de la siguiente forma

$$\begin{array}{c}
 \text{Tiempo } k \\
 \begin{array}{c}
 1 \\
 2 \\
 \vdots \\
 n
 \end{array}
 \end{array}
 \begin{array}{c}
 \text{Tiempo } k+1 \\
 \begin{array}{cccc}
 1 & 2 & \cdots & n
 \end{array} \\
 \left[ \begin{array}{cccc}
 P_{1,1} & P_{1,2} & \cdots & P_{1,n} \\
 P_{2,1} & P_{2,2} & \cdots & P_{2,n} \\
 \vdots & \vdots & \ddots & \vdots \\
 P_{n,1} & P_{n,2} & \cdots & P_{n,n}
 \end{array} \right]
 \end{array}$$

es decir, si actualmente nos encontramos en el estado 2 en un tiempo  $k$ , para saber en que estado se encontrará en el tiempo  $k + 1$  solamente será necesario analizar el segundo renglón de la matriz de transición. Entonces, esta matriz de transición ó probabilidad  $\mathbf{P}$  cumple con las siguientes características

$$\begin{array}{l} \text{a) } \mathbf{P} \geq 0 \\ \text{b) } \mathbf{P}\mathbf{e}' = \mathbf{e}' \end{array} \quad (3.4)$$

que nos dice que los elementos de la matriz siempre serán positivos ó cero y la segunda condición nos dice que la suma por renglones es igual a 1, donde  $\mathbf{e}'$  es el vector columna de puros 1's. La segunda condición nos dice que la matriz de transición tiene como valor propio a  $\lambda = 1$  con vector propio asociado  $\mathbf{e}'$ , y más adelante esta propiedad se verá reflejada en información importante que se puede obtener de la matriz de transición. Si uno está interesado en saber cuál es la probabilidad de ir del estado  $i$  al estado  $j$  en dos o más pasos, es importante considerar las ecuaciones de Chapman-Kolmogorov<sup>1</sup>, que dan una forma muy sencilla de obtener estas probabilidades en forma matricial cuando el proceso es homogéneo.

**Proposición 3.5** *Para una cadena de Markov  $\{X_k\}_{k \in \mathbb{N}}$  con espacio de estados  $\mathbb{E}$  y  $\forall n, m \in \mathbb{N}$ ,  $i, j \in \mathbb{E}$  se cumple*

$$P_{i,j}^{(n+m)} = P[X_{n+m} = j \mid X_0 = i] = \sum_{k \in \mathbb{E}} P_{i,k}^{(n)} P_{k,j}^{(m)} \quad (3.5)$$

*Demostración.* La demostración sólo utiliza los conceptos de función marginal y de probabilidad condicional

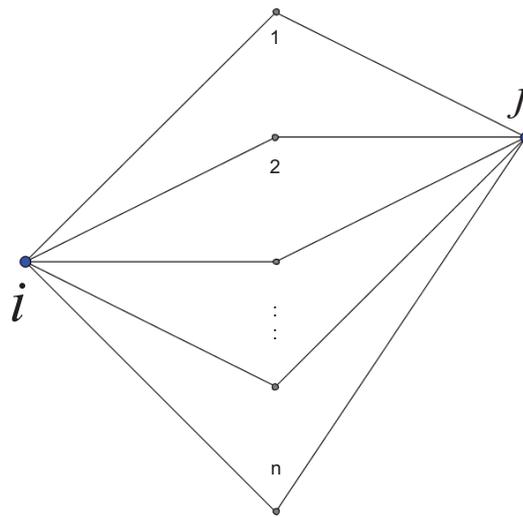
$$\begin{aligned} P_{i,j}^{(n+m)} &= \sum_{k \in \mathbb{E}} P[X_{n+m} = j, X_n = k \mid X_0 = i] \quad \text{si } 0 < n < n + m \\ &= \sum_{k \in \mathbb{E}} \frac{P[X_{n+m} = j, X_n = k, X_0 = i]}{P[X_0 = i]} \\ &= \sum_{k \in \mathbb{E}} P[X_{n+m} = j \mid X_n = k, X_0 = i] P[X_n = k \mid X_0 = i] \\ &= \sum_{k \in \mathbb{E}} P[X_{n+m} = j \mid X_n = k] P[X_n = k \mid X_0 = i] \\ &= \sum_{k \in \mathbb{E}} P_{k,j}^{(m)} P_{i,k}^{(n)} = \sum_{k \in \mathbb{E}} P_{i,k}^{(n)} P_{k,j}^{(m)} \end{aligned}$$

■

Esta relación implica que si tenemos un espacio de estados  $\mathbb{E}$  finito entonces esta relación se puede escribir en términos matriciales como  $\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)}\mathbf{P}^{(m)}$ , es decir, solamente será necesario conocer las probabilidades de transición en  $n$  y  $m$  pasos respectivamente, multiplicar las matrices, y así poder obtener las probabilidades en  $n + m$  pasos, por lo que hay que

<sup>1</sup>Sydney Chapman (1888-1970). Matemático inglés. Trabajó también en las áreas de astronomía y geofísica. En el área que logró tener mayor desarrollo fue en la de los procesos estocásticos. Las ecuaciones que llevan su nombre fueron desarrollados independientemente por él y por Kolmogorov.

recorrer todos los pasos intermedios. Un diagrama que ayuda a entender con mayor claridad la ecuación (3.5) se muestra a continuación.



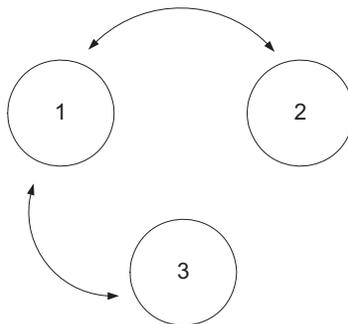
Con las ecuaciones de Chapman-Kolmogorov se “barren” todos los estados intermedios que conforman el espacio de estados

Falta encontrar las probabilidades de transición en  $n$  pasos, ya que la fórmula (3.5) no nos dice como calcular estas probabilidades de manera práctica, y con las ecuaciones de Chapman-Kolmogorov se obtienen estas probabilidades de manera matricial.

**Teorema 3.6 (Chapman-Kolmogorov)** *Sea  $\{X_k\}_{k \in \mathbb{E}}$  una cadena de Markov con espacio de estados finito  $\mathbb{E}$  y con matriz de transición  $\mathbf{P}$ . Entonces las probabilidades de transición en  $n$  pasos son la  $n$ -ésima potencia de la matriz de transición*

$$\mathbf{P}^{(n)} = \mathbf{P}^n \tag{3.6}$$

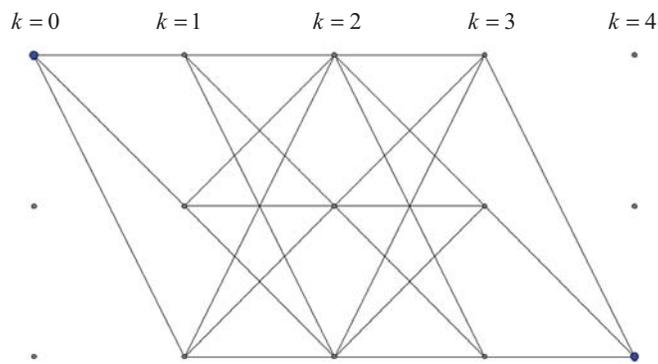
A continuación se analizará un ejemplo de estas probabilidades. Supongamos que tenemos un proceso estocástico  $\{X_n\}_{n \in \mathbb{N}}$  con un espacio de estados  $\mathbb{E} = \{1, 2, 3\}$  y cuyo diagrama de transiciones se muestra a continuación.



las probabilidades que se pueden asignar a esta forma de transición entre los estados es

$$\mathbf{P} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

con esto se obtienen las probabilidades en un paso del proceso. Por el momento no es importante la asignación de las probabilidades de las transiciones, inicialmente se tomarán probabilidades de manera uniforme, ya que solo es necesario saber como da "el salto" el proceso en cada instante del tiempo. Para saber la probabilidad  $P_{1,3}^{(4)} = P[X_4 = 3 | X_0 = 1]$ , que nos indica cuál es la probabilidad de que el estado se encuentre en 3 en cuatro pasos dado que empezó en 1, si no tuvieramos la relación (3.6) tendríamos que visualizar todas las posibles trayectorias, como se muestra a continuación:



algunas de estas trayectorias valen 0, porque no puede haber transición del estado 1 al 1 en un paso. Pero al tener las ecuaciones de C-K nos permiten obtener fácilmente esta probabilidad con una simple multiplicación de matrices, y en Mat/Oct es fácil de calcular.

```
>> P=[0 0.5 0.5; 1 0 0; 1 0 0]
```

```
P =
```

```

      0    0.5000    0.5000
  1.0000      0      0
  1.0000      0      0
```

```
>> % Calculamos las probs de trans de 4 pasos
```

```
>> P4 = P^4
```

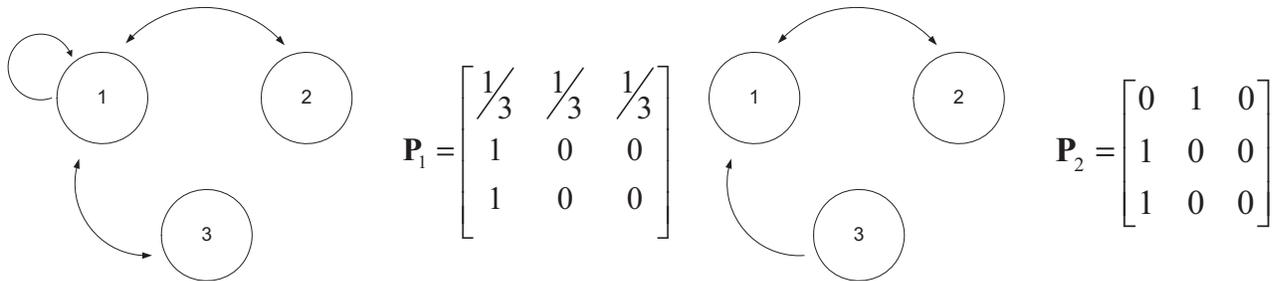
```
P4 =
```

```

  1.0000      0      0
      0    0.5000    0.5000
      0    0.5000    0.5000
```

Con esto observamos que  $P_{1,3}^{(4)} = 0$ , es decir, que no existe ninguna trayectoria que nos permita llegar del estado 1 al estado 3 en cuatro pasos. También la entrada (1, 1) de la matriz  $\mathbf{P}^{(4)}$  nos dice que si el proceso inicia en el estado 1, con probabilidad 1 estará el proceso de

nuevo en este estado. Al cambiar ligeramente las transiciones entre los estados se obtienen diferentes matrices de transición y se puede diferenciar entre categorías de estados, como se observa en los siguientes diagramas:



Analizando de igual manera la probabilidad  $P_{1,3}^{(4)}$  en estos dos casos obtenemos los siguientes resultados

```
>> % Analizamos el proceso P1
>> P1=[1/3 1/3 1/3; 1 0 0; 1 0 0]
P1 =
    0.3333    0.3333    0.3333
    1.0000         0         0
    1.0000         0         0
```

```
>> P14=P1^4
P14 =
    0.6790    0.1605    0.1605
    0.4815    0.2593    0.2593
    0.4815    0.2593    0.2593
```

```
>> % Analizamos el proceso P2
>> P2=[0 1 0; 1 0 0; 1 0 0]
P2 =
     0     1     0
     1     0     0
     1     0     0
```

```
>> P24=P2^4
P24 =
     1     0     0
     0     1     0
     0     1     0
```

Para el primer proceso con matriz de transición  $\mathbf{P}_1$  se obtiene que  $P_{1,3}^{(4)} = 0.1605$ , es decir, existe un 16.05% de probabilidad de que si el proceso inicio en el estado 1 termine en el estado 3 en cuatro pasos. Para el proceso con matriz de transición  $\mathbf{P}_2$  se tiene que  $P_{1,3}^{(4)} = 0$  y observando el diagrama de salto, si el proceso inicia en el estado 3 inmediatamente sale

de este estado y el proceso jamás regresará a ese estado, por lo que se puede asegurar que  $P_{1,3}^{(n)} = 0$  y  $P_{2,3}^{(n)} = 0$  para  $n \in \mathbb{N}$ , en este caso se dice que el estado 3 es transitorio. Con este ejemplo se observa que en una cadena de Markov un estado  $i \in \mathbb{E}$  se puede clasificar en dos tipos, puede ser un **estado recurrente** o un **estado transitorio**. La idea intuitiva de un estado recurrente es que el proceso lo podrá visitar una infinidad de veces y en el otro caso, un estado es transitorio si el proceso sólo lo visita una cantidad finita de veces y en algún momento el proceso estocástico ya no regresará más a este estado, lo abandona para siempre. Para definir con mayor claridad el concepto de estados recurrentes y transitorios, se usa el concepto de comunicación entre dos estados.

**Definición 3.7** Sean  $i, j \in \mathbb{E}$ . Se dice que el estado  $i$  se **comunica** con el estado  $j$ , que se denota como  $i \rightarrow j$ , si es posible en un número finito de pasos pasar del estado  $i$  al  $j$  ( $\exists n_0 \in \mathbb{N} \ni P_{i,j}^{(n_0)} > 0$ ). Si el estado  $i$  no se comunica con el estado  $j$ , lo denotaremos como  $i \nrightarrow j$ . Si  $i$  se comunica con  $j$ , y  $j$  se comunica con  $i$ , entonces decimos que los estados  $i$  y  $j$  se **comunican** y lo denotaremos como  $i \leftrightarrow j$ . De manera similar, si  $i$  no se comunica con  $j$  y/o  $j$  no se comunica con  $i$  lo denotaremos como  $i \nleftrightarrow j$ .

**Definición 3.8** Un estado  $i \in \mathbb{E}$  se dice que es **recurrente** si  $\forall j \in \mathbb{E}$  tal que  $i \rightarrow j \implies j \rightarrow i$ . Un estado  $i \in \mathbb{E}$  se dice que es **transitorio** si no es recurrente ( $\exists j \in \mathbb{E}$  tal que  $i \rightarrow j$  pero  $j \nrightarrow i$ )

**Definición 3.9** Para un estado  $i \in \mathbb{E}$  se definen como la **probabilidad de primer arribo** al tiempo  $n \in \mathbb{N}$  a la expresión

$$f_{i,i}^{(n)} = P[X_n = i, X_{n-1} \neq i, X_{n-2} \neq i, \dots, X_1 \neq i \mid X_0 = i] \quad (3.7)$$

y la **probabilidad de un eventual retorno** al estado  $i \in \mathbb{E}$  está dada por

$$f_{i,i} = \sum_{k=0}^{\infty} f_{i,i}^{(k)}$$

En Norris [23] se demuestra como intervienen las probabilidades de primer arribo y de transición para clasificar a un estado en recurrente o transitorio.

**Teorema 3.10** Sea  $\{X_n\}_{n \in \mathbb{N}}$  una cadena de Markov con espacios de estados finito  $\mathbb{E}$  y sea  $f_{i,i}$  las probabilidad de un eventual retorno

$$\begin{aligned} & a) \text{ Si } f_{i,i} = 1 \text{ entonces } i \in \mathbb{E} \text{ es } \mathbf{recurrente} \text{ y } \sum_{n=1}^{\infty} P_{i,i}^{(n)} = \infty \\ & b) \text{ Si } f_{i,i} < 1 \text{ entonces } i \in \mathbb{E} \text{ es } \mathbf{transitorio} \text{ y } \sum_{n=1}^{\infty} P_{i,i}^{(n)} < \infty \end{aligned} \quad (3.8)$$

Clasificar correctamente los estados de una matriz de transición asociada a una cadena de Markov en recurrentes o transitorios es un problema importante, debido a que mucha de la información vital que se puede extraer de un proceso depende de si el proceso solo tiene estados recurrentes o por lo menos existe un estado transitorio. En Xie y Beerel [24] se muestra una forma más práctica de poder analizar los estados en recurrentes y transitorios de manera eficiente, y para ello construyen dos conjuntos denotados como el **forward set** y el **backward set**. El conjunto *forward set* consiste en todos los estados con los se comunica  $i_0$ , es decir  $\mathcal{F}(i_0) = \{j \in \mathbb{E} \mid i_0 \rightarrow j\}$  y el conjunto *backward set* es el conjunto de todos los estados que se comunican con  $i_0$ , es decir  $\mathcal{B}(i_0) = \{j \in \mathbb{E} \mid j \rightarrow i_0\}$ . Todo el procedimiento se basa en ir analizando por renglón de la matriz  $\mathbf{P}$  las entradas positivas. La forma en que se plantea el algoritmo en el artículo es un poco difícil de entender, pero en este caso al realizar el código en Mat/Oct para ir encontrando el conjunto *forward set* se buscan las entradas positivas de los renglones de  $\mathbf{P}$  y en el caso del *backward set* se buscan las entradas positivas de los renglones de la matriz transpuesta  $\mathbf{P}^T$ . Esto genera una pequeña diferencia en relación a como se explica en el artículo, esto se debe a que Mat/Oct puede encontrar entradas de la matriz satisfagan una característica en particular, usando el comando **find**. En Xie y Beerel [24] vienen los elementos necesarios para poder clasificar un estado en recurrente o transitorio, aquí solo se expondrán los elementos generales.

**Lema 3.11** Sean  $i, j \in \mathbb{E}$ . Si  $j \in \mathcal{F}(i)$  entonces  $\mathcal{F}(j) \subseteq \mathcal{F}(i)$ . De manera similar si  $j \in \mathcal{B}(i)$  entonces  $\mathcal{B}(j) \subseteq \mathcal{B}(i)$

**Teorema 3.12** El estado  $i \in \mathbb{E}$  es **recurrente**  $\Leftrightarrow \mathcal{F}(i) \subseteq \mathcal{B}(i)$ . En otras palabras,  $i \in \mathbb{E}$  es **transitorio**  $\Leftrightarrow \mathcal{F}(i) \not\subseteq \mathcal{B}(i)$ .

**Teorema 3.13** Si el estado  $i \in \mathbb{E}$  es transitorio, entonces todos los estados en  $\mathcal{B}(i)$  son transitorios. Si  $i \in \mathbb{E}$  es recurrente, los estados en  $\mathcal{F}(i)$  son recurrentes, y el conjunto  $\mathcal{B}(i) \setminus \mathcal{F}(i)$  (si no es vacío) contiene solamente estados transitorios.

Se construyen los siguientes archivos m-file con una implementación del algoritmo de clasificación de estados, al cuál llamaremos *analisis\_estados.m*.

```
function [recurrentes,transitorios,clas_rec] = analisis_estados(P)

% ESTA FUNCION REGRESA LOS ESTADOS RECURRENTES Y TRANSITORIOS
% DE LA MATRIZ DE TRANSICION "P" USANDO UN ALGORITMO SIMILAR
% DESCRITO EN EL ARTICULO "Efficient State Clasiffication of
% Finite-State Markov Chains" DE AIGUO XIE Y PETER A. BEEREL.
% EN "clas_rec" SE REGRESA CUANTOS ESTADOS PERTENECEN A CADA
% CLASE DE COMUNICACION DE LOS ESTADOS RECURRENTES.
%
% FORMA DE USARSE: " [recurrentes,transitorios,clas_rec] = analisis_estados(P) "
```

```
% CALCULAMOS LA DIMENSION DE "P"
[n1,n2]=size(P);
```

```

% PARA PODER ENCONTRAR EL BACKWARD SET
Pt=P';

% POR SI NO REPRESENTA A UNA CADENA DE MARKOV
if (n1~=n2)
    error(' LA MATRIZ NO ES CUADRADA!!! ')
end

% TOMAMOS LA DIMENSION DE LA MATRIZ
n=max([n1 n2]);

% CREAMOS VARIABLES AUXILIARES
temporal = 1:n; % EL ESPACIO DE ESTADOS TEMPORAL
% PARA SABER QUE ESTADOS SON RECURRENTE Y CUALES SON TRANSITORIOS
recurrentes=[]; transitorios=[]; clas_rec=[];

while( ~isempty(temporal) )
    k =temporal(round(unifrnd(1,length(temporal)))));
    % TOMAMOS UN ELEMENTO ALEATORIAMENTE
    Fk = forward_set(P,k); % ENCONTRAMOS EL FORWARD SET
    Bk = forward_set(Pt,k); % ENCONTRAMOS EL BACKWARD SET
    if( all(ismember(Fk,Bk)) == 1 ) % EL ESTADO "k" ES RECURRENTE
        recurrentes=[recurrentes Fk];
        % UNIMOS LOS ESTADOS RECURRENTE
        clas_rec=[clas_rec length(Fk)];
        % PARA SABER CUANTOS ELEMENTOS PERTENECEN A CADA CLASE
        if(~isempty(setdiff(Bk,Fk))) % LOS OTROS ESTADOS SON TRANSITORIOS
            transitorios = union(transitorios,setdiff(Bk,Fk));
        end
    else
        transitorios = union(transitorios,union(k,Bk));
        % EL ESTADO "k" ES TRANSITORIO, Y TAMBIEN LO ES B(k)
    end
    temporal = setdiff(temporal,union(k,Bk));
    % BORRAMOS LOS ESTADOS DEL BACKWARD SET PARA REDUCIR EL ESPACIO DE ESTADOS
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function Fj=forward_set(P,j)

% ESTA FUNCION REGRESA EL CONJUNTO "FORWARD SET"
% DEL ESTADO "j" CON MATRIZ DE TRANSICION "P".
% EN OTRAS PALABRAS
%
```

```

%          Fj = { i | j -> i }
%
% FORMA DE USARSE: " Fj = forward_set(P,j) "

% VEMOS CON QUE ESTADOS SE COMUNICA INICIALMENTE EL ESTADO "j"
F0 = find(P(j,:));

% "F1" Y "F2" SON VECTORES AUXILIARES
F1 = []; F2 = setdiff(F0,F1);

% EN "F1" GUARDAMOS LOS ESTADOS CON LOS QUE ESTADO COMUNICADO "j" Y EN
% "F2" ANALIZAMOS LOS ESTADOS QUE NO SE HAN COMUNICADO CON "j"
while( ~isempty(F2) )
    for k=F2
        F1=union(F1,find(P(k,:)));
    end
    F2 = setdiff(F1,F0);
    F0 = union(F0,F1);
end

% SE ENCUENTRA EL CONJUNTO FORWARD SET
Fj = F0;

```

A continuación se muestran dos ejemplos que ejemplifican el uso de este código que será de mucha utilidad al querer descomponer los estados en recurrentes y transitorios de una cadena de Markov. Primero tomaremos la siguiente matriz de transición de algún proceso  $\{X_n\}_{n \in \mathbb{N}}$  con espacio de estados  $\mathbb{E} = \{1, 2, 3, \dots, 9\}$

$$\mathbf{P} = \begin{bmatrix} 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Como  $\mathbf{P}$  se puede considerar una matriz "grande" no es tan sencillo encontrar el diagrama de transiciones, por lo que no se ve de manera inmediata si todos los estados son recurrentes o existe al menos un estado transitorio. Para saber esto usaremos el código y veremos como se clasifican los estados de este proceso.

```

>> P=[0 0.5 0 0 0.5 0 0 0 0; 0 0 1 0 0 0 0 0 0 ; 0 0 0 1 0 0 0 0 0 ;
    1 0 0 0 0 0 0 0 0; 0 0 0 0 0 1 0 0 0; 0 0 0 0 0 0 1 0 0;
    0 0 0 0 0 0 0 1 0; 0 0 0 0 0 0 0 0 1; 1 0 0 0 0 0 0 0 0];

```

```
>> [recurrentes,transitorios,clas_rec]= analisis_estados(P)
recurrentes =
    1     2     3     4     5     6     7     8     9

transitorios =
    []

clas_rec =
    9
```

Estos resultados nos revelan que todos los estados del proceso son recurrentes, no hay estados transitorios y lo más importante es que todos los 9 estados recurrentes pertenecen a una sola de comunicación, por lo que se dice que todo el proceso es **irreducible**. Adelante se dará una definición más formal sobre clases de comunicación. Otra matriz de transición con nueve estados está dada por

$$\mathbf{P}_1 = \begin{bmatrix} 0 & 0 & 0 & 0.2 & 0 & 0 & 0 & 0 & 0.8 \\ 0 & 0.1 & 0.3 & 0 & 0.5 & 0 & 0 & 0 & 0.1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 0 & 0.4 & 0.1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0.4 \end{bmatrix} \tag{3.9}$$

al igual que en el caso anterior, es una matriz "grande", y no se pueden visualizar rápidamente las transiciones y el análisis de que estados son recurrentes o transitorios no es tan sencillo. Utilizando el código obtenemos los siguientes resultados:

```
>> P1=[0 0 0 0.2 0 0 0 0 0.8; 0 0.1 0.3 0 0.5 0 0 0 0.1;
    0 0 0 0 0 0 0 1 0; 1 0 0 0 0 0 0 0 0; 0 0 0 0 1 0 0 0 0;
    0 1 0 0 0 0 0 0 0; 0 0.5 0 0 0 0.4 0.1 0 0;
    0 0 1 0 0 0 0 0 0; 0 0 0 0.6 0 0 0 0 0.4];

>> [recurrentes,transitorios,clas_rec]= analisis_estados(P1)
recurrentes =
    3     8     1     4     9     5

transitorios =
    2     6     7

clas_rec =
    2     3     1
```

Esto nos dice que este proceso está compuesto de estados recurrentes y transitorios, además dentro de los estados recurrentes existen varias clases de comunicación. La primera

clase de comunicación de los estados recurrentes está compuesta por 2 estados  $R_1 = \{3, 8\}$ , la segunda clase de comunicación está compuesta por 3 estados  $R_2 = \{1, 4, 9\}$  y la última clase de comunicación está compuesta por un solo estado  $R_3 = \{5\}$  (a este tipo de estados que solamente consisten en un solo estado se les conoce como **estados absorbentes**, ya que una vez que el proceso entra a este estado jamás se saldrá de él). Por último los estados transitorios están en  $T = \{2, 6, 7\}$ , por lo que se tiene que el espacio de estados es la unión de estos 4 conjuntos  $\mathbb{E} = R_1 \cup R_2 \cup R_3 \cup T$  y si reacomodamos los renglones y columnas de la matriz en un nuevo orden

$$\mathbb{E}^* = \{3, 8, 1, 4, 9, 5, 2, 6, 7\}$$

obtendremos esta nueva matriz de transición donde se observará con mayor claridad la estructura matricial inducida por las clases de comunicación.

$$\mathbf{P}_1^* = \left[ \begin{array}{cc|ccc|ccc} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0.2 & 0.8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.6 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \hline 0.3 & 0 & 0 & 0 & 0.1 & 0.5 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0.4 & 0.1 \end{array} \right] = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix}, \quad (3.10)$$

la submatriz  $\mathbf{C}$  contiene las distintas clases de comunicación de los estados recurrentes, la submatriz de ceros  $\mathbf{0}$  es porque no existe comunicación de los estados recurrentes a los estados transitorios. La submatriz  $\mathbf{Q}$  contiene las probabilidades de transición entre los estados transitorios  $\mathbf{Q} = (P_{i,j})_{i,j \in T}$  y por último la submatriz  $\mathbf{R}$  contiene las probabilidades de transición de los estados transitorios a los estados recurrentes  $\mathbf{R} = (P_{i,j})_{i \in T, j \in R}$ . Cuando se puede transformar una matriz a esta forma se dice que el proceso es **reducible**. A continuación se dará una definición formal a los elementos que se han visto hasta el momento.

**Definición 3.14** Sea  $i \in \mathbb{E}$ . La **clase de comunicación** del estado  $i$  se define como:

$$C_i = \{j \in \mathbb{E} \mid i \leftrightarrow j\} \quad (3.11)$$

**Definición 3.15** Sea  $i \in \mathbb{E}$ . Si  $\forall i \in \mathbb{E}$  tenemos que  $C_i = \mathbb{E}$ , decimos que la matriz de transición asociada a la cadena de Markov es **irreducible**. De lo contrario, si existen  $k$  clases de comunicación ( $k > 1$ ) con  $C_i \cap C_j = \emptyset$  para  $i \neq j$  y  $\bigcup_k C_k = \mathbb{E}$  entonces decimos que la matriz de transición asociada a la cadena de Markov es **reducible**.

Para dar una definición más matricial de cuando un proceso es reducible o no, es necesario considerar permutaciones de un arreglo numérico  $\{1, 2, \dots, n\}$

**Definición 3.16** Un conjunto de permutaciones  $\sigma_j$  del conjunto  $\{1, 2, \dots, n\}$  en si misma es denotada por  $S_n$ . Un elemento de  $S_n$  esta definido por

$$\sigma_1 = i_1, \sigma_2 = i_2, \dots, \sigma_n = i_n \tag{3.12}$$

donde  $i_1, i_2, \dots, i_n$  es algún reordenamiento de los elementos de  $\{1, 2, \dots, n\}$ . Esta permutación se escribe de la forma

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma_1 & \sigma_2 & \cdots & \sigma_n \end{pmatrix} \tag{3.13}$$

**Definición 3.17** Una **matriz de permutación**<sup>2</sup>  $\mathbf{S} \in \mathbb{R}^{n \times n}$  consiste en un arreglo  $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_n\}$  de los elementos  $\{1, 2, \dots, n\}$  donde cada uno de los elementos de  $\mathbf{S}$  están conformados de la siguiente manera:

$$S_{i,j} = \delta_{i,\sigma_j} = \begin{cases} 1 & \text{si } i = \sigma_j \\ 0 & \text{si } i \neq \sigma_j \end{cases} \tag{3.14}$$

**Definición 3.18** Una matriz  $\mathbf{P} \in \mathbb{R}^{n \times n}$  de probabilidad se dice que es **reducible** si existe una matriz de permutación  $\mathbf{S} \in \mathbb{R}^{n \times n}$  con la propiedad:

$$\mathbf{S}^T \mathbf{P} \mathbf{S} = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix} \tag{3.15}$$

donde la matriz  $\mathbf{0}$  está conformada por puros 0's y las matrices  $\mathbf{C}$  y  $\mathbf{Q}$  son matrices cuadradas. Si no existe tal matriz de permutación diremos que la matriz  $\mathbf{P}$  es **irreducible**.

Se puede ver que un proceso es irreducible si todos los estados dentro del espacio muestral se comunican entre sí, y por lo tanto todos son estados recurrentes. En cambio, un proceso es reducible si por lo menos existen dos clases de comunicación recurrentes y/o al menos un estado transitorio. Si un proceso es reducible, es fácil poner a la matriz de transición en su forma reducible, solamente hay que acomodar el arreglo de permutación  $\sigma$  con la siguiente estructura:

$$\sigma = \{\text{recurrentes}, \text{transitorios}\}$$

dentro de los estados recurrentes acomodarlos conforme a sus clases de comunicación  $R = R_1 \cup R_2 \cup \dots \cup R_m$ . Se puede construir fácilmente un archivo m-file para obtener la matriz de permutación.

---

<sup>2</sup>En la literatura una matriz de permutación se denota como  $\mathbf{P}$ , pero para evitar confusiones entre una matriz de permutación y una matriz de probabilidad a la matriz de permutación se le denotará con la letra  $\mathbf{S}$

```

function S=permutacion(sigma)

% ESTA FUNCION REGRESA LA MATRIZ DE PERMUTACION
% "T" ASOCIADA AL ARREGLO "sigma", QUE DEBE SER
% ALGUNA PERMUTACION.
%
% FORMA DE USARSE: " S = permutacion(sigma) "

% PARA SABER LA DIMENSION DEL ARREGLO
n = length(sigma);

% TIENE QUE SER UNA PERMUTACION ENTRE 1 Y n
sigma = abs(sigma);
sigma = round(sigma);

% HAY ALGUN ERROR EN LA PERMUTACION
if( ~isempty(find(sigma==0 | sigma>n)) || ~isequal(sort(sigma),1:n) )
    error(' LA PERMUTACION INTRODUCIDA NO ES VALIDA!!! ')
end

% CREAMOS NUESTRA MATRIZ DE PERMUTACION
S=zeros(n);
for i=1:n
    for j=1:n
        if( i==sigma(j) )
            S(i,j)=1;
        else
            S(i,j)=0;
        end
    end
end
end

```

A continuación se mostrarán los resultados básicos sobre cada tipo de proceso.

### 3.1.1 Cadenas irreducibles

Una de las características básicas de los procesos irreducibles es conocer como se comunican los estados entre sí. A este concepto se le conoce como **periodicidad**.

**Definición 3.19** *Se dice que el estado  $i \in \mathbb{E}$  tiene **periodo**  $d$  si ocurre lo siguiente:*

$$a) P_{i,i}^{(n)} = 0 \text{ a menos que } n = md \text{ con } m \in \mathbb{N}$$

b)  $d$  es el mayor entero que cumple esto.

(3.16)

**Proposición 3.20** *El periodo  $d_i$  de un estado  $i \in \mathbb{E}$  es igual a*

$$d_i = \text{mcd}\{n > 0 \mid P_{i,i}^{(n)} > 0\} \quad (3.17)$$

donde  $\text{mcd}$  denota el **máximo común divisor** del conjunto  $d_i$ .

**Definición 3.21** *Una proceso irreducible es **aperiódico** si sus estados tienen periodo 1.*

Hay toda una teoría especializada cuando se tiene que un proceso irreducible tiene periodo  $d \geq 2$ , para ello se recomienda consultar las referencias Meyer [1], Gantmacher [40] y [23]. Un resultado práctico (que no en todos los casos sucede) es que si el proceso es irreducible y algún elemento de la diagonal es positivo, entonces todo el proceso es automáticamente aperiódico. De manera práctica si se tiene cualquier estado  $j$  y  $P_{j,j} > 0$ , entonces por la ecuación (3.17) se cumple que

$$\begin{aligned} d_j &= \text{mcd} \left\{ n \geq 1 \mid P_{j,j}^{(n)} > 0 \right\} \\ &= \text{mcd} \left\{ 1, \text{mcd} \left\{ n \geq 2 \mid P_{j,j}^{(n)} > 0 \right\} \right\} \\ &= 1 \end{aligned}$$

y con esto se obtiene que todo el proceso es aperiódico. Aunque esto no quiere decir que si los elementos de la diagonal sean cero el proceso es periódico, es una condición necesaria pero no suficiente. Por ejemplo, si se tiene la matriz de un proceso irreducible con  $\mathbb{E} = \{1, 2, 3, 4\}$  dada por

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (3.18)$$

este proceso es aperiódico. Solo basta analizar el periodo del estado 3

$$d_3 = \text{mcd} \{2, 3, 4, 5, \dots\} = 1$$

por lo tanto todo el proceso es aperiódico. Un resultado matricial para garantizar que un proceso irreducible es aperiódico es que dada la matriz de transición  $\mathbf{P} \in \mathbb{R}^{n \times n}$  entonces se cumple  $\mathbf{P}^{(n^2-2n+2)} > \mathbf{0}$ , aunque para matrices con un espacio de estados pequeño sería fácil verificar este resultado, para matrices muy grandes sería un trabajo computacionalmente complicado. Una de las ventajas de tener un proceso aperiódico es que el límite de la matriz existe y es único. Para el caso de la matriz anterior (3.18) se tiene que el límite es

$$\lim_{n \rightarrow \infty} \mathbf{P}^{(n)} = \begin{bmatrix} 0.2 & 0.2 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 \end{bmatrix} = \mathbf{\Pi} \quad (3.19)$$

Este resultado se interpreta como que el proceso visita al estado 1 el 20% de las veces, visita al estado 3 el 40% de las veces y así respectivamente. Encontrar esta matriz es fundamental dentro de las cadenas de Markov, por lo que se enunciarán algunos resultados sobre este tema.

**Teorema 3.22** Sea  $\{X_k\}_{k \in \mathbb{N}}$  una cadena de Markov irreducible con espacio de estados finito y matriz de transición  $\mathbf{P}$ . Si el proceso es aperiódico entonces existe una distribución de probabilidad  $\{\pi_j\}_{j \in \mathbb{E}}$  y números  $\alpha > 0$  y  $\beta \in (0, 1)$  tales que  $\forall i, j \in \mathbb{E}$

$$\left| P_{i,j}^{(k)} - \pi_j \right| \leq \alpha \beta^k \quad \text{para } k = 1, 2, \dots \quad (3.20)$$

Este resultado nos indica que existe un vector de probabilidad  $\boldsymbol{\pi} = [\pi_1 \ \pi_2 \ \dots \ \pi_n]$  al que las potencias convergen, independientemente del estado que se haya iniciado. Una posible manera de demostrar este resultado es pensando que la matriz de transición es diagonalizable, es decir, que todos sus valores propios son distintos. Si tenemos que la matriz de transición  $\mathbf{P} \in \mathbb{R}^{n \times n}$  se puede descomponer como  $\mathbf{P} = \mathbf{BDB}^{-1}$  donde  $\mathbf{D}$  es la matriz diagonal con los valores propios  $\mathbf{D} = \text{diag}(1, \lambda_2, \dots, \lambda_n)$ ,  $\mathbf{B}$  es la matriz cuyas columnas son los vectores propios de  $\mathbf{P}$  entonces

$$\begin{aligned} \mathbf{P} &= \mathbf{BDB}^{-1} = \begin{bmatrix} \mathbf{e}' & \mathbf{x}'_2 & \dots & \mathbf{x}'_n \end{bmatrix} \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{bmatrix} \\ &= \mathbf{e}'\mathbf{y}_1 + \lambda_2\mathbf{x}'_2\mathbf{y}_2 + \dots + \lambda_n\mathbf{x}'_n\mathbf{y}_n \end{aligned} \quad (3.21)$$

a esta forma de construir a la matriz de transición se le conoce como la **descomposición espectral**. La primera columna de la matriz  $\mathbf{B}$  es el vector de 1's debido a la segunda condición de (3.4). La ventaja de tener la descomposición espectral es que las probabilidades de transición se calculan fácilmente:

$$\mathbf{P}^{(k)} = \mathbf{P}^k = \mathbf{e}'\mathbf{y}_1 + (\lambda_2)^k \mathbf{x}'_2\mathbf{y}_2 + \dots + (\lambda_n)^k \mathbf{x}'_n\mathbf{y}_n$$

con esto se puede verificar que  $\lim_{k \rightarrow \infty} \mathbf{P}^{(k)} = \boldsymbol{\Pi} = \mathbf{e}'\mathbf{y}_1$  y utilizando las propiedades de las normas matriciales se tiene:

$$\begin{aligned} \|\mathbf{P}^{(k)} - \boldsymbol{\Pi}\| &= \|(\lambda_2)^k \mathbf{x}'_2\mathbf{y}_2 + \dots + (\lambda_n)^k \mathbf{x}'_n\mathbf{y}_n\| \\ &\leq |\lambda_2|^k \|\mathbf{x}'_2\mathbf{y}_2\| + \dots + |\lambda_n|^k \|\mathbf{x}'_n\mathbf{y}_n\| \end{aligned}$$

y si tomamos  $\alpha = \sum_{k \neq 1} \|\mathbf{x}'_k\mathbf{y}_k\|$  y  $\beta = \max_{k \neq 1} |\lambda_k|$ , entonces

$$\|\mathbf{P}^{(k)} - \boldsymbol{\Pi}\| \leq \alpha \beta^k \quad \text{con } \alpha \geq 0, \beta \in [0, 1) \quad (3.22)$$

que es una versión matricial de la ecuación (3.20). La ecuación anterior nos da la idea de que la convergencia a la matriz límite depende mucho del segundo valor propio más grande de la matriz, entre más cercano sea a 1 entonces más lenta será la convergencia. El problema central es poder encontrar de manera "fácil" la matriz límite  $\boldsymbol{\Pi}$  y para ello hay algunos

resultados importantes, que se enuncian sin demostración. Se dice que una matriz  $\mathbf{A} \in \mathbb{R}^{n \times n}$  no negativa es **regular (o primitiva)** si existe un  $n_0 \in \mathbb{N}$  tal que  $\mathbf{A}^{n_0} > \mathbf{0}$ . Para el resultado demostrado por Perron<sup>3</sup> y Frobenius<sup>4</sup> se utiliza el concepto de **matriz adjunta**, que a su vez necesita de la matriz de cofactores. Si denotamos a la matriz de cofactores de una matriz cualquiera  $\mathbf{A}$  por  $\mathbf{C} = \text{cof}(\mathbf{A})$ , entonces la matriz adjunta está definida de la siguiente manera:

$$\text{adj}(\mathbf{A}) = \begin{cases} \mathbf{A} \text{ singular} \Rightarrow \text{adj}(\mathbf{A}) = \mathbf{C}^T \\ \mathbf{A} \text{ no singular} \Rightarrow \text{adj}(\mathbf{A}) = \det(\mathbf{A}) \cdot \mathbf{A}^{-1} \end{cases} \quad (3.23)$$

**Teorema 3.23 (Perron-Frobenius)** *Sea  $\mathbf{A} \in \mathbb{R}^{n \times n}$  una matriz regular. Entonces existe un valor propio  $\lambda_1 > 0$  que excede los valores absolutos de los demás valores propios de  $\mathbf{A}$ . Al valor propio  $\lambda_1$  se le pueden asociar vectores propios por la izquierda y por la derecha  $\mathbf{y}_1$  y  $\mathbf{v}'_1$  que son únicos salvo por constantes. Si  $\mathbf{y}_1$  y  $\mathbf{v}'_1$  se seleccionan tal que  $\mathbf{y}_1 \mathbf{v}'_1 = 1$  entonces*

$$\mathbf{v}'_1 \mathbf{y}_1 = \frac{\text{adj}(\lambda_1 \mathbf{I} - \mathbf{A})}{p'(\lambda_1)} \quad (3.24)$$

donde  $p(\lambda)$  es el polinomio característico de  $\mathbf{A}$  y  $p'(\lambda)$  es la derivada de  $p(\lambda)$  con respecto a  $\lambda$ .

**Teorema 3.24 (Perron-Frobenius 2)** *Sea  $\{X_k\}_{k \in \mathbb{N}}$  una cadena de Markov finita, irreducible y aperiódica con matriz de transición  $\mathbf{P}$ . Entonces la matriz límite  $\mathbf{\Pi}$  está dada por*

$$\mathbf{\Pi} = \lim_{n \rightarrow \infty} \mathbf{P}^{(n)} = \mathbf{e}' \mathbf{y}_1 = \frac{\text{adj}(\mathbf{I} - \mathbf{P})}{p'(1)} \quad (3.25)$$

donde  $p(\lambda)$  es el polinomio característico de  $\mathbf{P}$  y  $p'(\lambda)$  es la derivada de  $p(\lambda)$  con respecto a  $\lambda$ .

En Gantmacher [40] e Iturrarán [28] se especifica que no es necesario calcular toda la matriz adjunta  $\text{adj}(\mathbf{I} - \mathbf{P})$  y que sólo es necesario calcular cualquier renglón de esta matriz adjunta. El problema es que si la matriz es de dimensión  $n$  entonces calcular este vector límite  $\boldsymbol{\pi}$  tiene un costo computacional de aproximadamente  $\frac{2}{3}n^4$  operaciones, por lo que para matrices de gran tamaño no es un método práctico. A continuación se muestra un código en Mat/Oct para el cálculo de este vector límite, usando el algoritmo descrito en Iturrarán [28].

<sup>3</sup>Oskar Perron (1880-1975). Matemático alemán. Hizo contribuciones importantes en el área de las ecuaciones diferenciales y teoría de matrices.

<sup>4</sup>Ferdinand Georg Frobenius (1849-1917). Matemático alemán. Hizo contribuciones importantes en el área de teoría de grupos, así como ayudó a formalizar la teoría de matrices como una rama formal de las matemáticas. Fue él quién al final de su carrera que trabajó con las matrices no negativas y definió en el concepto de irreducibilidad de una matriz.

```

function xlim = vector_limite(P)

% ESTA FUNCION REGRESA EL VECTOR LIMITE "xlim"
% DE UN PROCESO IRREDUCIBLE Y APERIODICO CON
% MATRIZ DE TRANSICION "P" USANDO EL METODO DE
% LA MATRIZ ADJUNTA.
%
% FORMA DE USARSE: xlim = vector_limite(P)

[n,m]=size(P); % Verificamos la dimension
if( n~=m )
    error(' LA MATRIZ "P" DEBE SER CUADRADA!!! ')
end

xlim=zeros(1,n); % Guardamos espacio en la memoria
A=eye(n)-P;
% Borrarnos primer columna
A(:,1)=[];

%Calculamos el primer renglon de la matriz adjunta
for j=1:n
    B=A; % Guardamos una "copia" de la matriz original
    B(j,:)=[]; % Borrarnos el j-esimo renglon
    xlim(j)=power(-1,j+1)*det(B); % Calculamos el cofactor j de B
end

% Dividimos entre la derivada del polinomio caracteristico en lam=1
xlim=xlim/polyval(polyder(poly(P)),1);

```

A continuación se muestra un ejemplo de como funciona el código usando como ejemplo la matriz de transición (3.18) y la matriz límite (3.19) obtenida haciendo cálculos sucesivos.

```

>> P = [0 1 0 0; 0 0 1 0; 0.5 0 0 0.5; 0 0 1 0]
P =
     0     1.0000         0         0
     0         0     1.0000         0
    0.5000         0         0     0.5000
     0         0     1.0000         0

>> xlim = vector_limite(P)
xlim =
    0.2000    0.2000    0.4000    0.2000

```

Con este procedimiento se observa que es un método bueno poder encontrar el vector límite en términos de la matriz adjunta, aunque en la práctica calcular el polinomio carac-

terístico de la matriz  $\mathbf{I} - \mathbf{P}$  es un problema serio, debido al tiempo de almacenamiento de memoria y el cálculo de demasiados determinantes. Es por esta razón que solo se recomienda que se utilice para matrices de tamaño pequeño, porque el cálculo del polinomio característico puede tener complicaciones numéricas para matrices de gran tamaño. Una forma más fácil para encontrar esta distribución límite se muestra a continuación.

**Lema 3.25** *Si la cadena de Markov  $\{X_k\}_{k \in \mathbb{N}}$  es irreducible y aperiódica con matriz de transición  $\mathbf{P} \in \mathbb{R}^{n \times n}$ , entonces el vector límite  $\boldsymbol{\pi}$  es la única solución al sistema de ecuaciones lineales*

$$\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{P} \quad \text{con } \boldsymbol{\pi} \mathbf{e}' = 1 \tag{3.26}$$

La condición (3.26) también implica que  $\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{P}^{(n)}$  para  $n = 1, 2, \dots$  por lo que al vector límite  $\boldsymbol{\pi}$  también se le conoce como el **vector estacionario**. Una de las características importantes es que el vector  $\boldsymbol{\pi}$  es un vector de probabilidad y sus elementos suman 1. Manipulando la ecuación  $\boldsymbol{\pi} = \boldsymbol{\pi} \mathbf{P}$  obtenemos que  $\boldsymbol{\pi}(\mathbf{I} - \mathbf{P}) = \mathbf{0}$ , usando la transpuesta se obtiene el siguiente sistema de ecuaciones lineales  $(\mathbf{I} - \mathbf{P})^T \boldsymbol{\pi}' = \mathbf{0}'$ , pero antes de poder resolver este sistema de ecuaciones se debe tener en cuenta que este sistema de ecuaciones no es de rango completo.

El **rango** de una matriz  $\mathbf{A} \in \mathbb{R}^{n \times m}$  se define como el número máximo de renglones (o columnas) linealmente independientes de  $\mathbf{A}$ , al cual se denota por  $R_{\mathbf{A}}$ . El **kernel** de una matriz está dado por  $K_{\mathbf{A}} = \{\mathbf{x}' \in \mathbb{R}^m \mid \mathbf{A} \mathbf{x}' = \mathbf{0}'\}$  y la nulidad de una matriz es la dimensión del kernel  $N_{\mathbf{A}} = \dim K_{\mathbf{A}}$ . Uno de los resultados importantes dentro del álgebra lineal es el siguiente

$$R_{\mathbf{A}} + N_{\mathbf{A}} = n \tag{3.27}$$

Se sabe que un sistema de ecuaciones  $\mathbf{A} \mathbf{x}' = \mathbf{b}'$  con  $\mathbf{A} \in \mathbb{R}^{n \times n}$  tiene una única solución si el rango de la matriz es completo, es decir,  $R_{\mathbf{A}} = n$  y  $N_{\mathbf{A}} = 0$ .

Para encontrar el vector estacionario  $\boldsymbol{\pi}$  se tiene que resolver el sistema de ecuaciones  $(\mathbf{I} - \mathbf{P})^T \boldsymbol{\pi}' = \mathbf{0}'$  pero el rango de la matriz  $\mathbf{I} - \mathbf{P}^T$  es  $n - 1$ , esto se debe a que  $N_{\mathbf{I} - \mathbf{P}^T} = \dim(\{\mathbf{x}' \in \mathbb{R}^n \mid (\mathbf{I} - \mathbf{P}^T) \mathbf{x}' = \mathbf{0}'\}) = \dim(\{\mathbf{x}' \in \mathbb{R}^n \mid \mathbf{x}' = \mathbf{P}^T \mathbf{x}'\}) = 1$ , y por lo tanto se obtiene que  $R_{\mathbf{I} - \mathbf{P}^T} = n - 1$ , esto implica que una de las ecuaciones de  $(\mathbf{I} - \mathbf{P})^T \boldsymbol{\pi}' = \mathbf{0}'$  resulta redundante, y se puede sustituir por la condición  $\mathbf{e} \boldsymbol{\pi}' = 1$ . Regresando al proceso (3.18), si se desea encontrar el vector límite y estacionario  $\boldsymbol{\pi}$ , primero se encuentra la matriz  $\mathbf{I} - \mathbf{P}^T$

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad \mathbf{I} - \mathbf{P}^T = \begin{bmatrix} 1 & 0 & -0.5 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & -0.5 & 1 \end{bmatrix}$$

y para encontrar  $\boldsymbol{\pi}$  inicialmente se debería de resolver el sistema de ecuaciones

$$\begin{bmatrix} 1 & 0 & -0.5 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & -0.5 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

pero por la implicación de que  $R_{\mathbf{I}-\mathbf{P}^T} = 3$  no hay una forma directa de resolver este sistema y además  $\sum \pi_j = 1$ . Entonces se pueden sustituir cualquier ecuación e introducir la restricción  $\sum \pi_j = 1$ . En este caso se considera que se puede sustituir cualquier ecuación, por ejemplo la primera ecuación o la última ecuación.

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 \\ 0 & 0 & -0.5 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{ó} \quad \begin{bmatrix} 1 & 0 & -0.5 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

En ambos casos se obtiene como solución  $\boldsymbol{\pi} = [0.2 \ 0.2 \ 0.4 \ 0.2]$ . Inicialmente no hay diferencia si se modifica el primer o el último renglón. En Meyer [29] se analiza como se altera el vector límite al hacer pequeñas perturbaciones a la matriz de transición. Para ello construye la siguiente matriz especial  $\mathbf{P}_\varepsilon \in \mathbb{R}^{n \times n}$  dada por

$$\mathbf{P}_\varepsilon = \begin{bmatrix} 1 - \varepsilon & \varepsilon/(n-1) & \cdots & \varepsilon/(n-1) \\ \varepsilon/(n-1) & 1 - \varepsilon & \cdots & \varepsilon/(n-1) \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon/(n-1) & \varepsilon/(n-1) & \cdots & 1 - \varepsilon \end{bmatrix}$$

donde  $\varepsilon$  es una cantidad tan pequeña como queramos y también tiene la particularidad de que es una matriz doblemente estocástica, estas matrices tienen la particularidad de que tanto los renglones como las columnas suman 1 y por lo tanto su distribución estacionaria es la distribución uniforme, dada por  $\boldsymbol{\pi} = \frac{1}{n}\mathbf{e}$ . Se han hecho algunos experimentos en Mat/Oct (se vieron limitados por la capacidad de memoria de la PC) para verificar que tan eficiente es modificar el primer renglón o el último renglón con esta matriz  $\mathbf{P}_\varepsilon$ . El método 1 se refiere a que se modificó el último renglón de la matriz y para el método 2 se modificó el primer renglón de la matriz. Se muestran algunos resultados obtenidos, donde se tomó el valor de  $\varepsilon = 2.22044604925031 \times 10^{-16}$ , que es el epsilon de la máquina con el que trabaja Mat/Oct, y que en magnitud es más o menos equivalente a  $\frac{1}{5000000000000000}$  y que en términos prácticos es una probabilidad "despreciable". Se muestra el primer y último elemento de cada método, debido a que los valores intermedios si resultaron ser iguales. Se utilizó el formato de doble

precisión.

$n$	Método 1		Método 2	
	$\pi_1^{(1)}$	$\pi_n^{(1)}$	$\pi_1^{(2)}$	$\pi_n^{(2)}$
2	0.5	0.5	0.5	0.5
10	0.1	0.09999999999999998	0.1	0.1
50	0.02000000000000002	0.01999999999999993	0.01999999999999994	0.02
100	0.009999999999999736	0.01000000000000019	0.00999999999999925	0.01
1000	0.000999999999983146	0.00100000000001599	0.00099999999999133	0.001
10000	0.00010000000016687	$9.9999999269711 \times 10^{-5}$	0.000100000000093818	0.0001

Estos resultados preliminares nos muestran que el método 1 no es tan eficiente, debido a que los valores obtenidos de la primera y última entrada difieren mucho del valor exacto. De manera empírica se usará el método 2 de modificar el primer renglón de preferencia, y de ser posible verificar la magnitud de las entradas de la matriz irreducible. En Stewart [13] se hace un análisis exhaustivo de los métodos convenientes para resolver el sistema de ecuaciones general  $(\mathbf{I} - \mathbf{P}^T)\boldsymbol{\pi}' = \mathbf{0}'$ , donde se muestran comparativos de que tipo de método de resolución es más efectivo, se ilustra que con los métodos iterativos se obtienen mejores resultados debido a que no se modifica la matriz original y se mantiene la estructura dispersa de una matriz. En nuestro caso sólo utilizaremos el método de sustituir el primer renglón para encontrar el vector límite y estacionario. Como conclusión se puede tener la siguiente tabla básica.

- a)  $\mathbf{P}$  es irreducible y aperiódico  $\implies \boldsymbol{\pi}$  es el vector límite y estacionario
- b)  $\mathbf{P}$  es irreducible y de periodo  $d \implies \boldsymbol{\pi}$  es el vector promedio y estacionario

Referencias clásicas del tema son Kemeny y Snell [30], Isaacson y Madsen[52] o una más reciente se puede encontrar en Bremaud [31]. A continuación se mostrarán de manera concreta algunos resultados básicos de los procesos reducibles.

### 3.1.2 Cadenas Reducibles

En el caso de que se trabaje con un proceso reducible, pueden existir 2 escenarios posibles. El primer caso para procesos reducibles es que no haya estados transitorios y existen por lo menos dos clases de comunicación de estados recurrentes. El segundo escenario es que haya al menos una clase de comunicación de estados recurrentes y por lo menos un estado transitorio. Para el primer caso es importante destacar que si existen al menos dos clases de estados recurrentes se puede pensar que se tienen varias matrices de transición dentro de una matriz de transición más grande, y con ello se puede obtener este primer resultado.

**Teorema 3.26** Sean  $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_k$  matrices de transición finitas irreducibles con vectores estacionarios  $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots, \boldsymbol{\pi}_k$  y sea  $\mathbf{P} = \text{diag}(\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_k)$  una matriz de transición formada por bloques diagonales. Si se tiene una combinación lineal de constantes no negativas

$\lambda_1, \lambda_2, \dots, \lambda_k$  tales que  $\sum \lambda_j = 1$  entonces el vector

$$\boldsymbol{\pi}^* = [ \lambda_1 \boldsymbol{\pi}_1 \quad \lambda_2 \boldsymbol{\pi}_2 \quad \cdots \quad \lambda_k \boldsymbol{\pi}_k ]$$

es un vector estacionario de  $\mathbf{P} = \text{diag}(\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_k)$

Este resultado implica que una matriz con al menos dos clases recurrentes tendrá una infinidad de vectores estacionarios. A continuación se muestra un ejemplo de este importante resultado dentro de los procesos reducibles.

```
>> P1=[0.3 0.7; 0.5 0.5]; P2=[0 1 0; 0 0 1; 1 0 0];
>> P=blkdiag(P1,P2)
P =
    0.3000    0.7000         0         0         0
    0.5000    0.5000         0         0         0
         0         0         0    1.0000         0
         0         0         0         0    1.0000
         0         0    1.0000         0         0

>> n1=vector_estacionario(P1)
n1 =
    0.4167    0.5833

>> n2=vector_estacionario(P2)
n2 =
    0.3333    0.3333    0.3333

>> L1=rand(1); L2=1-L1; N=[L1*n1 L2*n2]; Mat=[N; N*P]
Mat =
    0.3806    0.5328    0.0289    0.0289    0.0289
    0.3806    0.5328    0.0289    0.0289    0.0289

>> L1=rand(1); L2=1-L1; N=[L1*n1 L2*n2]; Mat=[N; N*P]
Mat =
    0.0406    0.0569    0.3008    0.3008    0.3008
    0.0406    0.0569    0.3008    0.3008    0.3008

>> L1=rand(1); L2=1-L1; N=[L1*n1 L2*n2]; Mat=[N; N*P]
Mat =
    0.2279    0.3190    0.1510    0.1510    0.1510
    0.2279    0.3190    0.1510    0.1510    0.1510
```

Observamos que un proceso reducible tiene una infinidad de vectores estacionarios si hay al menos dos clases recurrentes. Otra propiedad importante de los procesos reducibles es

cuando hay al menos un estado transitorio se puede escribir a la matriz de transición  $\mathbf{P}$  en su forma canónica o reducible usando (3.15)

$$\mathbf{S}^T \mathbf{P} \mathbf{S} = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix} \quad (3.28)$$

donde la submatriz  $\mathbf{Q}$  contiene las probabilidades de transición entre estados transitorios. Como existe un momento donde ya el proceso deja de visitar a los estados transitorios, entonces se cumple que  $\lim_{n \rightarrow \infty} P_{i,j}^{(n)} = 0$  para  $i, j \in T$  y en términos matriciales esto nos dice que  $\lim_{n \rightarrow \infty} \mathbf{Q}^n = \mathbf{0}$ . Para este caso, surge una pregunta importante ¿en que momento el proceso deja de visitar a los estados transitorios?. Para verificar esto se define a la siguiente variable aleatoria

$$Y_{i,j}^{(k)} = \begin{cases} 1 & \text{si } X_k = j \text{ dado que } X_0 = i \\ 0 & \text{e.o.c} \end{cases}$$

entonces esta variable aleatoria registra si el proceso visitó al estado  $j$  en el  $k$ -ésimo paso dado que inició en el estado  $i$ . Lo que se desea es saber cuántas veces se visitaron estos estados en  $n$  pasos, por lo que se tiene la suma  $N_{i,j}^{(n)} = \sum_{k=1}^n Y_{i,j}^{(k)}$ , que a su vez es una variable aleatoria. Queremos saber en promedio cuántas veces se visitó el estado  $j$  en  $n$  pasos, se tiene

$$E \left[ N_{i,j}^{(n)} \right] = E \left[ \sum_{k=1}^n Y_{i,j}^{(k)} \right] = \sum_{k=1}^n E \left[ Y_{i,j}^{(k)} \right] = \sum_{k=1}^n P_{i,j}^{(k)}$$

y si denotamos a la matriz  $\mathbf{M}^{(n)} = \left( E \left[ N_{i,j}^{(n)} \right] \right)_{i,j}$  se obtiene el siguiente resultado en términos de las potencias de la matriz de transición

$$\mathbf{M}^{(n)} = \mathbf{P} + \mathbf{P}^{(2)} + \dots + \mathbf{P}^{(n)} \quad (3.29)$$

Con esto se obtiene lo que se conoce en la literatura como la **matriz fundamental**, en el contexto de los procesos estocásticos con matriz cadenas reducible.

**Teorema 3.27** *Para una cadena de Markov reducible la matriz  $\mathbf{I} - \mathbf{Q}$  tiene inversa y será denotada por  $\mathbf{M}$*

$$\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots \quad (3.30)$$

La entrada  $M_{i,j}$  representa el número esperado de veces que se visitó el estado  $j$  transitorio, dado que la cadena empezó en el estado  $i$  transitorio. El estado inicial se cuenta si  $i = j$ .

**Demostración.** Para ello tenemos la matriz (3.28) y la matriz de promedios de visitas (3.29). Tomando la multiplicación por bloques se obtiene

$$\mathbf{P}^{(n)} = \begin{bmatrix} \mathbf{C}^{(n)} & \mathbf{0} \\ * & \mathbf{Q}^n \end{bmatrix}$$

entonces tenemos que el número promedio de visitas a la larga está dada por  $\mathbf{M} = \lim_{n \rightarrow \infty} \mathbf{M}^{(n)}$  y por lo tanto

$$\mathbf{M} = \begin{bmatrix} \infty & \mathbf{0} \\ * & \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots \end{bmatrix}$$

esto porque las clases recurrentes se visitan una infinidad de veces. El número promedio de visitas entre los estados transitorios entonces están dados (abusando de la notación, ya que tomaremos sólo la submatriz) por  $\mathbf{M} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots$ . Como eventualmente el proceso abandonará a los estados transitorios, se sabe que  $\lim_{n \rightarrow \infty} P_{i,j}^{(n)} = 0$  con  $i, j \in T$  y matricialmente quedaría representado por

$$\lim_{n \rightarrow \infty} \mathbf{Q}^n = \mathbf{0}$$

por lo que se cumple que la matriz es convergente y el radio espectral de la matriz  $\mathbf{Q}$  es menor que 1, utilizando el resultado (2.5) y la fórmula clásica de convergencia de una serie de potencias  $\sum_{j=0}^{\infty} x^j = \frac{1}{1-x}$  con  $|x| < 1$  se obtiene el resultado buscado

$$\mathbf{M} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots = (\mathbf{I} - \mathbf{Q})^{-1}$$

■

Este resultado es muy importante, porque nos da una idea de cuánto tiempo el proceso estará dentro de la clase de estados transitorios antes de ser absorbido por una clase recurrente. Si queremos tomar sólo en cuenta cual es el tiempo promedio que le toma al estado  $k \in T$  salir de los estados transitorios tomamos únicamente la suma por renglones de la matriz  $\mathbf{M}$ , a la cuál denotaremos por  $\mathbf{M}^*$

$$\boxed{\mathbf{M}^* = \mathbf{M}\mathbf{e}'}$$
 (3.31)

Consideremos la matriz (3.10) y recordemos que el espacio de estados se actualizó a  $\mathbb{E}^* = \{3, 8, 1, 4, 9, 5, 2, 6, 7\}$ , se había encontrado que

$$\mathbf{Q} = \begin{bmatrix} 0.1 & 0 & 0 \\ 1 & 0 & 0 \\ 0.5 & 0.4 & 0.1 \end{bmatrix}$$
 (3.32)

Con esta matriz fundamental podremos encontrar el tiempo esperado antes de ser absorbido por algún estado recurrente.

$$\mathbf{M} = (\mathbf{I} - \mathbf{Q})^{-1} = \begin{bmatrix} 1.1111 & 0 & 0 \\ 1.1111 & 1 & 0 \\ 1.1111 & 0.4444 & 1.1111 \end{bmatrix}$$
 (3.33)
$$\mathbf{M}^* = \mathbf{M}\mathbf{e}' = \begin{bmatrix} 1.1111 \\ 2.1111 \\ 2.6666 \end{bmatrix}$$

Esto nos dice que por ejemplo, si el proceso empieza en el estado 7 transitorio, entonces aproximadamente en 2.6666 pasos saldrá de los estados transitorios para entrar a una clase recurrente. Si se empieza en el estado 6 se tardará aprox. 2.1111 pasos en salir de los estados transitorios y si se empieza en el estado 2 se tardará 1.1111 pasos en salir de los estados transitorios. Es importante destacar que los promedios de tiempos de absorción son solo estimaciones. Por último, se estudiarán las probabilidades de absorción, que son otra de las características de las cadenas reducibles. Queremos encontrar una respuesta a la pregunta ¿cuál será el estado recurrente que tiene mayor probabilidad de entrar el proceso cuando deja abandona la clase de estados transitorios?. Por el momento los estados recurrentes se transformarán en estados absorbentes, ya que se asume que cuando el proceso entra en alguno de ellos ya no se podrá abandonar este estado. Con esto se ve el siguiente resultado:

**Teorema 3.28** Sea  $\{X_k\}_{k \in \mathbb{N}}$  una cadena de Markov de un proceso reducible. Definamos a  $F_{i,j}$  como la probabilidad de que un estado transitorio  $i \in T$  sea absorbido por un estado absorbente  $j \in R$ . Sea  $\mathbf{F}$  la matriz con entradas  $f_{i,j}$  entonces  $\mathbf{F}$  está dada por

$$\boxed{\mathbf{F} = \mathbf{MR}} \tag{3.34}$$

*Demostración.* Basta analizar a la matriz en forma reducible dada por

$$\tilde{\mathbf{P}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix}$$

e ir tomando potencias sucesivas por bloques

$$\tilde{\mathbf{P}}^{(2)} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R} + \mathbf{QR} & \mathbf{Q}^2 \end{bmatrix} \qquad \tilde{\mathbf{P}}^{(3)} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{R} + \mathbf{QR} + \mathbf{Q}^2\mathbf{R} & \mathbf{Q}^3 \end{bmatrix}$$

$$\tilde{\mathbf{P}}^{(n)} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \left( \sum_{k=0}^{n-1} \mathbf{Q}^k \right) \mathbf{R} & \mathbf{Q}^n \end{bmatrix}$$

tomando la probabilidad a la larga se obtiene que

$$\lim_{n \rightarrow \infty} \tilde{\mathbf{P}}^{(n)} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ (\mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots) \mathbf{R} & \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{R} & \mathbf{0} \end{bmatrix}$$

con esto se obtiene el resultado deseado. ■

Como es muy importante estas dos matrices  $\mathbf{R}$  y  $\mathbf{Q}$  dentro de los procesos reducibles, es necesario que se pueda construir un archivo que permita la extracción fácil de estas dos matrices, a la cuál llamaremos *extraccionRQ.m*, y se muestra a continuación.

```
function [P,R,Q]=extraccionRQ(P)
```

```
% ESTA FUNCION EXTRAER LAS MATRICES "R" Y "Q" DE UN PROCESO
```

```

% REDUCIBLE CON AL MENOS UN ESTADO TRANSITORIO CON MATRIZ
% DE TRANSICION "P". PARA HACER EL ANALISIS SE UTILIZARA
% LA FUNCION "analisis_estados.m". LA FUNCION REGRESA A LA
% MATRIZ "P" EN SU FORMA REDUCIBLE.
%
% FORMA DE UTILIZARSE [P,R,Q]=extraccionRQ(P)

[n,m]=size(P);
if(n~=m)
    error(' La matriz no es cuadrada!!! ')
end

% Se analiza la matriz
[rec,trans,clas_rec]=analisis_estados(P);
Rn=length(rec); Tn=length(trans); Cn=length(clas_rec);

% Se analiza los tipos de formas que puede
% tener la matriz
if( Cn==1 && Tn==0 )
    error(' El proceso es irreducible!!! ')
elseif(Cn>1 && Tn==0)
    disp(' El proceso tiene varias clases recurrentes')
    error(' y ningun estado transitorio!!! ')
end

% Encontramos la permutacion sigma = [Rec Tran] y la matriz
% en su forma reducible y la extraccion de matrices
sigma = [ rec trans ]; P = P(sigma,sigma);
ext=(Rn+1):n; R=P(ext,1:Rn); Q=P(ext,ext);

```

Para ver como funciona este código, tomaremos la matriz (3.9) y su forma reducible en (3.10).

```

>> P1=[0 0 0 0.2 0 0 0 0 0.8; ...; 0 0 0 0.6 0 0 0 0 0.4];
>> [P1,R,Q]=extraccionRQ(P1);
>> R, Q
R =
    0    0    0.1000    0.3000    0    0.5000
    0    0         0         0         0         0
    0    0         0         0         0         0

Q =
    0.1000         0         0
    1.0000         0         0
    0.5000    0.4000    0.1000

```

Tenemos los elementos más importantes para las cadenas de Markov finitas tanto irreducibles como reducibles. Muchas de estas propiedades tienen su similitud al trabajar con los procesos de Markov finitos, y sólo será necesario analizar la estructura matemática y matricial de este tipo de procesos.

### 3.2 Procesos de Markov Finitos

A continuación se darán algunas propiedades básicas de los procesos de Markov, sus similitudes con las cadenas de Markov y algunos resultados importantes.

**Definición 3.29** Se dice que el proceso  $\{X(t)\}_{t \in \mathbb{R}^+}$  es un **proceso de Markov a tiempo continuo** con espacio de estados finito  $\mathbb{E}$  si para cada pareja de números reales  $s, t \geq 0$  e  $i, j \in \mathbb{E}$  se cumple la propiedad de Markov

$$P[X(t+s) = j \mid X(s) = i, X(u) : 0 \leq u < s] = P[X(t+s) = j \mid X(s) = i]. \quad (3.35)$$

Se dice que el proceso es **homogéneo o estacionario** si se cumple

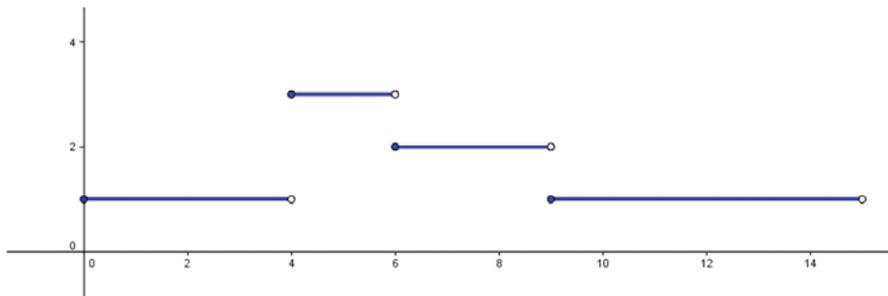
$$P[X(t+s) = j \mid X(s) = i] = P[X(t) = j \mid X(0) = i] = P_{i,j}(t).$$

Esta definición es una extensión natural de las cadenas de Markov a tiempo discreto. La matriz con probabilidades  $P_{i,j}(t)$  se denotará por  $\mathbf{P}(t)$ . En Kulkarni [34] se demuestra que también la matriz de transición satisface las propiedades

$$\begin{aligned} \text{a) } & \mathbf{P}(t) \geq \mathbf{0} \\ \text{b) } & \mathbf{P}(t)\mathbf{e}' = \mathbf{e}' \text{ para todo } t \in \mathbb{R}^+ \\ \text{c) } & \mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s) \text{ para } t, s \in \mathbb{R}^+ \\ \text{d) } & \lim_{t \rightarrow 0^+} \mathbf{P}(t) = \mathbf{I}. \end{aligned} \quad (3.36)$$

que son equivalentes a las de un proceso discreto, y las ecuaciones de Chapman-Kolmogorov están dadas en la propiedad c). Las últimas dos propiedades indican que las probabilidades de transición  $P_{i,j}(t)$  son continuas. Hasta aquí todo es similar a las matrices de transición a tiempo discreto, pero una de las características fundamentales de un proceso a tiempo continuo con espacio de estados finito es que hay una matriz implícita  $\mathbf{T}$  que define las "tasas de transición" entre estados. Esto nos dice que a tiempo discreto lo que define a un proceso son sus probabilidades de transición y a tiempo continuo son las tasas de transición. Además, al trabajar con esta matriz de tasas  $\mathbf{T}$  se pueden describir propiedades importantes

del proceso sin tener que estar trabajando con matrices de transición  $\mathbf{P}(t)$  para cada tiempo  $t$  que se desee analizar. La dinámica de un proceso continuo se describe a continuación, aunque la demostración formal no se mostrará, pero en Cinlar [35] viene todo el proceso de construcción sobre este proceso estocástico. Si el proceso inicia en el estado  $i \in \mathbb{E}$ , entonces el proceso se mantiene en ese estado siguiendo una distribución exponencial del parámetro  $T_i$ , es decir  $\mathcal{T}_i \sim \text{Exp}(T_i)$ , con  $\mathcal{T}_i$  el tiempo que se mantuvo el proceso en el estado  $i$ . Después el proceso realizará una transición a un estado  $j$  (con  $j \neq i$ ) con una probabilidad  $J_{i,j}$ . Con ello el proceso se queda un tiempo indeterminado  $\mathcal{T}_j$  en el estado  $j$  siguiendo una distribución exponencial de parámetro  $T_j$  y con ello realizar una transición a otro estado  $k$  con probabilidad  $J_{j,k}$  y así sucesivamente. Con esto se dará una definición sobre estos parámetros especiales  $T_k$ .



Ejemplo de la trayectoria que sigue un proceso de Markov

**Definición 3.30** Para un estado  $k \in \mathbb{E}$  de un proceso de Markov finito, se tienen tres posibilidades

- a) Es un estado absorbente si  $T_k = 0$
  - b) Es un estado estable si  $0 < T_k < +\infty$
  - c) Es un estado instantáneo si  $T_k = +\infty$
- (3.37)

Con lo expuesto anteriormente y esta definición se construye de una manera más formal a la matriz de tasas  $\mathbf{T}$  y a la matriz de salto<sup>5</sup>  $\mathbf{J}$ . Esta última matriz nos dirá como se están comunicando los estados dentro del proceso. La cantidad  $T_{i,j} = T_i J_{i,j}$  para  $i \neq j$  se le conoce como la **tasa de transición** del estado  $i$  al estado  $j$ . Los elementos en la diagonal de la matriz  $\mathbf{T}$  están dados de la siguiente manera

$$T_{k,k} = -T_k = - \sum_{j \neq k} T_{k,j}$$

(3.38)

<sup>5</sup>En la literatura no existe una convención única para denotar a esta matriz de saltos del proceso, por lo que asumiremos en este trabajo la notación  $\mathbf{J}$  por "jump" (salto en inglés) y no confundirla con la matriz de permutación  $\mathbf{S}$  que se definió anteriormente.

Esto indica que los elementos de la diagonal son la suma por renglón de todas las tasas de transición con signo negativo, ya que por cuestiones técnicas se pide que la suma por renglones de la matriz  $\mathbf{T}$  sea cero, es decir,  $\mathbf{T}\mathbf{e}' = \mathbf{0}'$  y esta última relación también nos indica que  $\lambda = 0$  es un valor propio con vector propio  $\mathbf{e}'$ . La matriz de saltos  $\mathbf{J}$  se define de la siguiente manera

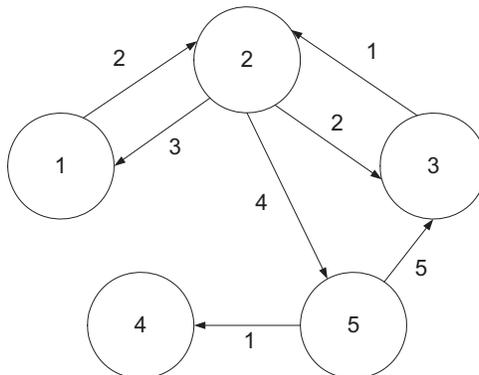
$$J_{i,j} = \begin{cases} T_{i,j}/T_i & \text{si } j \neq i \text{ y } T_i \neq 0 \\ 0 & \text{si } j \neq i \text{ y } T_i = 0 \end{cases} \quad J_{i,i} = \begin{cases} 0 & \text{si } T_i \neq 0 \\ 1 & \text{si } T_i = 0 \end{cases} \quad (3.39)$$

Se divide cada renglón por su elemento en la diagonal, si el renglón  $k$  es el vector de ceros entonces es un estado absorbente por lo que se pone un 1 en la diagonal, para los estados que no son absorbentes se pone un cero en la diagonal evitando transiciones hacia el mismo estado. La matriz de saltos  $\mathbf{J}$  nos permite conocer si un estado es recurrente o transitorio (como se hizo con los procesos discretos). La demostración de este resultado viene en Norris [23], y se enunciará a continuación.

**Teorema 3.31** *Sea  $\{X(t)\}_{t \in \mathbb{R}^+}$  un proceso de Markov finito con matriz de saltos  $\mathbf{J}$  asociado a un proceso discreto  $\{Y_n\}_{n \in \mathbb{N}}$ . Entonces se cumple que*

$$\begin{aligned} & a) \text{ Si el estado } i \text{ es recurrente (transitorio) en } \{Y_n\}_{n \in \mathbb{N}} \\ & \quad \text{entonces } i \text{ es recurrente (transitorio) en } \{X(t)\}_{t \in \mathbb{R}^+} \\ & b) \text{ Cualquier estado es recurrente ó transitorio} \\ & c) \text{ Recurrencia y transitoriedad son propiedades de clase} \end{aligned} \quad (3.40)$$

Este resultado es importante, ya que encontrar la matriz de saltos es sencillo y se puede utilizar el programa m-file *analisis\_estados.m* para clasificar de manera sencilla los estados en recurrentes y transitorios. En la teoría de procesos a tiempo continuo se debería de verificar que la siguiente integral  $\int_0^\infty P_{k,k}(t)dt$  converge o no, si converge es transitorio el estado  $k$  y si diverge entonces es recurrente, usar esta definición sería imposible de llevarla a la práctica para procesos con espacio de estados grande. A continuación se muestra un ejemplo con todos estos elementos. Supongamos que tenemos un proceso de Markov  $\{X(t)\}_{t \in \mathbb{R}^+}$  con espacio de estados  $\mathbb{E} = \{1, 2, 3, 4, 5\}$  cuyas tasas de transición se muestran sobre los arcos



Se puede pensar que es un proceso de Markov debido a que sólo es importante en que estado se encuentra antes de dar el salto a otro estado. La matriz de tasas "incompleta"  $\mathbf{T}_{\text{incomp}}$  y la "normal"  $\mathbf{T}$  está dada por

$$\mathbf{T}_{\text{incomp}} = \begin{bmatrix} 0 & 2 & 0 & 0 & 0 \\ 3 & 0 & 2 & 0 & 4 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 \end{bmatrix} \quad \mathbf{T} = \begin{bmatrix} -2 & 2 & 0 & 0 & 0 \\ 3 & -9 & 2 & 0 & 4 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & -4 \end{bmatrix} \quad (3.41)$$

con esto se observa que primero debemos encontrar las transiciones entre los estados y por último sumamos los renglones y la suma con signo invertido es el que se pone en la diagonal. Como el estado 4 no se comunica con ningún otro estado entonces este es un estado absorbente. Usando la propiedad 3.39 se obtiene que la matriz de saltos  $\mathbf{J}$  para este proceso está dada por

$$\mathbf{J} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 3/9 & 0 & 2/9 & 0 & 4/9 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 3/4 & 1/4 & 0 \end{bmatrix}$$

y con esto se puede analizar que los estados 1, 2, 3 y 5 son estados transitorios y el estado 4 es recurrente absorbente. También muestra este ejemplo que toda la información necesaria del proceso se puede obtener a partir de la matriz de tasas  $\mathbf{T}$ , con lo que será necesario revisar sus propiedades básicas. Las tasas  $T_{i,i}$  y  $T_{i,j}$  describen como es el movimiento del proceso para intervalos de tiempo pequeños

$$\begin{aligned} P[X(t+h) = j \mid X(t) = i] &= T_{i,j}h + o(h) \\ P[X(t+h) = i \mid X(t) = i] &= 1 - T_{i,i}h + o(h) \end{aligned} \quad (3.42)$$

donde la función  $o(h)$  se puede pensar como una "función despreciable", es decir

$$\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0 \quad (3.43)$$

y estas dos relaciones indican la "intensidad" con la que ocurre un cambio de estado en un intervalo corto de tiempo. En Rolski et. al [61] se demuestra que propiedades matemáticas cumple la matriz de tasas

$$\boxed{\lim_{h \rightarrow 0^+} \frac{P_{i,i}(h) - 1}{h} = -T_{i,i} \quad \lim_{h \rightarrow 0^+} \frac{P_{i,j}(h)}{h} = T_{i,j}} \quad (3.44)$$

y esta última relación en términos matriciales nos dice lo siguiente

$$\boxed{\left. \frac{d\mathbf{P}(t)}{dt} \right|_{t=0} = \lim_{h \rightarrow 0^+} \frac{\mathbf{P}(h) - \mathbf{I}}{h} = \mathbf{T}} \quad (3.45)$$

que es la intensidad con la que ocurre un cambio de estado al iniciar el proceso. Aquí lo importante es que aunque no se conocen de manera explícita las probabilidades  $\mathbf{P}(t)$  tenemos la información de como se comportan cuando el proceso inicia. Esta última relación (3.45) permitirá encontrar una forma explícita de como encontrar las probabilidades de transición a cualquier momento, y para ello veremos lo que se conoce en la literatura como las ecuaciones de Kolmogorov hacia adelante y hacia atrás.

**Teorema 3.32 (Kolmogorov)** *Si se tiene un proceso de Markov  $\{X(t)\}_{t \in \mathbb{R}^+}$  con espacio de estados  $\mathbb{E}$  y matriz de tasas  $\mathbf{T}$ , entonces se cumplen las siguientes relaciones*

$$\begin{aligned} a) \text{ Ecuaciones Kolmogorov hacia atrás } \frac{d\mathbf{P}(t)}{dt} &= \mathbf{T}\mathbf{P}(t) \\ b) \text{ Ecuaciones Kolmogorov hacia adelante } \frac{d\mathbf{P}(t)}{dt} &= \mathbf{P}(t)\mathbf{T} \end{aligned} \tag{3.46}$$

Para encontrar estas ecuaciones solamente se debe tomar las ecuaciones de Chapman-Kolmogorov  $\mathbf{P}(t+s) = \mathbf{P}(t)\mathbf{P}(s)$ , derivar con respecto a cada variable y evaluar  $t = 0$  ó  $s = 0$  según convenga. La importancia de estas ecuaciones es que en el caso finito es que se tiene una forma explícita para calcular las probabilidades de transición  $\mathbf{P}(t)$  para cada instante  $t$ , usando el resultado visto anteriormente en (2.9) y obtener que

$$\mathbf{P}(t) = e^{\mathbf{T}t} \text{ con } t \in [0, \infty) \tag{3.47}$$

recordando la definición (2.6) se puede verificar fácilmente que representa una matriz de transición para cada momento  $t$ .

$$\begin{aligned} \mathbf{P}(t)\mathbf{e}' &= \left( \mathbf{I} + t\mathbf{T} + \frac{t^2\mathbf{T}^2}{2} + \frac{t^3\mathbf{T}^3}{3!} + \dots \right) \mathbf{e}' \\ &= \mathbf{I}\mathbf{e}' + t\mathbf{T}\mathbf{e}' + \frac{t^2}{2}\mathbf{T}^2\mathbf{e}' + \frac{t^3}{3!}\mathbf{T}^3\mathbf{e}' + \dots = \mathbf{I}\mathbf{e}' = \mathbf{e}' \end{aligned}$$

esto se debe a la propiedad de que la suma por renglones de la matriz de tasas es cero  $\mathbf{T}\mathbf{e}' = \mathbf{0}'$  y por tanto  $\mathbf{T}^n\mathbf{e}' = \mathbf{0}'$  para  $n \in \mathbb{N}$ . Previamente se analizó los distintos métodos para poder encontrar de manera eficiente estas probabilidades  $\mathbf{P}(t) = e^{\mathbf{T}t}$ . A continuación se darán resultados básicos tanto de los procesos irreducibles como reducibles de este proceso y verificar que tiene una similitud con las cadenas de Markov finitas.

### 3.2.1 Procesos Irreducibles

Al trabajar con un proceso irreducible, se sabe que todos los estados están comunicados entre sí. La ventaja de los procesos de Markov es que no existe en sí el concepto de periodicidad,

por lo que el proceso es esencialmente aperiódico. Esto quiere decir que siempre existirá el siguiente límite

$$\lim_{t \rightarrow \infty} \mathbf{P}(t) = \mathbf{\Pi}$$

En Kijima [36] se demuestra que la convergencia al límite es de orden geométrico

$$\|\mathbf{P}(t) - \mathbf{\Pi}\|_{\infty} \leq C\tau(t)$$

donde  $\tau(\cdot)$  es el coeficiente de ergodicidad de la matriz<sup>6</sup>. La ventaja es que para encontrar esta matriz límite tenemos que realizar un procedimiento similar al que se hizo cuando se trabajó con las cadenas de Markov irreducibles.

**Definición 3.33** Se dice que el vector  $\boldsymbol{\pi} \in \mathbb{R}^{1 \times n}$  es una distribución estacionaria para el proceso irreducible finito  $\{X(t)\}_{t \in \mathbb{R}^+}$  si se cumple que para todo  $t \in \mathbb{R}^+$

$$\boxed{\boldsymbol{\pi} \mathbf{P}(t) = \boldsymbol{\pi}} \tag{3.48}$$

Para encontrar este vector estacionario, que a su vez cumplirá con  $\mathbf{\Pi} = \mathbf{e}'\boldsymbol{\pi}$ , solamente es necesario trabajar con la definición de la matriz exponencial

$$\begin{aligned} \boldsymbol{\pi} \mathbf{P}(t) &= \boldsymbol{\pi} \iff \boldsymbol{\pi} \left( \mathbf{I} + \mathbf{T}t + \frac{\mathbf{T}^2 t^2}{2} + \frac{\mathbf{T}^3 t^3}{3!} + \dots \right) = \boldsymbol{\pi} \\ \iff \boldsymbol{\pi} + \boldsymbol{\pi} \mathbf{T}t + \boldsymbol{\pi} \mathbf{T}^2 \frac{t^2}{2} + \boldsymbol{\pi} \mathbf{T}^3 \frac{t^3}{3!} + \dots &= \boldsymbol{\pi} \\ \iff \boldsymbol{\pi} \mathbf{T}t + \boldsymbol{\pi} \mathbf{T}^2 \frac{t^2}{2} + \boldsymbol{\pi} \mathbf{T}^3 \frac{t^3}{3!} + \dots &= \mathbf{0} \text{ para todo } t \\ \iff \boldsymbol{\pi} \mathbf{T}^k = \mathbf{0} \text{ para } k \geq 1 & \\ \iff \boldsymbol{\pi} \mathbf{T} = \mathbf{0} & \end{aligned}$$

esta última expresión nos dice que para encontrar el vector estacionario debemos encontrar el vector propio por la izquierda de la matriz de tasas  $\mathbf{T}$  asociado al valor propio  $\lambda = 0$  junto con la restricción de que sea un vector de probabilidad  $\boldsymbol{\pi} \mathbf{e}' = 1$ . Acomodándolo como un sistema de ecuaciones tendremos que resolver el siguiente sistema de ecuaciones

$$\boxed{\mathbf{T}^T \boldsymbol{\pi}' = \mathbf{0}' \text{ con } \boldsymbol{\pi} \mathbf{e}' = 1} \tag{3.49}$$

Entonces, como en el caso de las cadenas reducibles, para un proceso irreducible el rango de la matriz de tasas  $\mathbf{T}$  es  $n - 1$ , por lo que se puede sustituir cualquier ecuación con la

<sup>6</sup>El coeficiente de ergodicidad de una matriz de transición  $\mathbf{P}$  está dado por

$$\tau(\mathbf{P}) = \frac{1}{2} \max_{i,j \in \mathbb{E}} \sum_{k \in \mathbb{E}} |P_{i,k} - P_{j,k}|$$

condición de que la suma sea 1. De nueva cuenta se muestra la semejanza entre la cadena de Markov y el proceso de Markov finitos. Como ejemplo se propone la siguiente matriz de tasas de un proceso irreducible

$$\mathbf{T} = \begin{bmatrix} -40 & 35 & 2 & 3 \\ 1 & -1 & 0 & 0 \\ 1 & 1 & -3 & 1 \\ 4 & 5 & 6 & -15 \end{bmatrix}$$

para encontrar el vector estacionario debemos resolver el siguiente sistema de ecuaciones

$$\begin{bmatrix} -40 & 1 & 1 & 4 \\ 35 & -1 & 1 & 5 \\ 2 & 0 & -3 & 6 \\ 3 & 0 & 1 & -15 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

se puede sustituir cualquier ecuación con la restricción  $\boldsymbol{\pi}\mathbf{e}' = 1$  por lo que, como se hizo antes, se sustituirá el primer renglón para obtener

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 35 & -1 & 1 & 5 \\ 2 & 0 & -3 & 6 \\ 3 & 0 & 1 & -15 \end{bmatrix} \begin{bmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

dando como resultado numérico aproximado que

$$\boldsymbol{\pi} = [ 0.0249 \quad 0.9374 \quad 0.0307 \quad 0.007 ]$$

En Stewart [37] se expone que en particular no se debe aplicar un método directo para resolver este sistema de ecuaciones, debido a que al aplicar la eliminación Gaussiana común se pueden perder dígitos significativos por los elementos negativos de la diagonal al realizar las sustracciones correspondientes, por lo que se recomienda el algoritmo propuesto por los autores Grassmann, Taksar y Heyman que utiliza una versión más estable para poder encontrar el vector estacionario, ya que se evita hacer las restas y con esto no se pierde exactitud en los resultados. A continuación se muestra el código propuesto en [37] del algoritmo GTH

```
function xest = vector_estacionarioGTH(T)

% ESTA FUNCION REGRESA EL VECTOR ESTACIONARIO DE UN
% PROCESO DE MARKOV IRREDUCIBLE CON MATRIZ DE TASAS
% "T", USANDO EL ALGORITMO DE Grassmann--Taksar--Heyman
%
% FORMA DE USARSE: xest = vector_estacionarioCMTTC(T)

% Codigo descrito en el libro de William Stewart
% "PROBABILITY, MARKOV CHAINS, QUEUES, AND SIMULATION"
```

```

A = T';
n = size(A);
for i=1:n-1
    escala = sum(A(i+1:n,i));
    for k=i+1:n
        A(i,k) = A(i,k)/escala;
    end

    for j=i+1:n
        for k=i+1:n
            A(j,k) = A(j,k) + A(j,i)*A(i,k);
        end
    end
end

xest(n) = 1;
for i=n-1:-1:1
    for j=i+1:n
        xest(i) = xest(i) + A(i,j)*xest(j);
    end
end
xest = xest/norm(xest,1);

```

Con este código se puede encontrar de manera directa el vector estacionario. Considerando el ejemplo anterior

```

>> T=[-40 35 2 3; 1 -1 0 0; 1 1 -3 1; 4 5 6 -15]
T =
   -40    35     2     3
     1    -1     0     0
     1     1    -3     1
     4     5     6    -15

>> xest=vector_estacionarioGTH(T)
xest =
    0.0249    0.9374    0.0307    0.0070

```

Estos resultados nos indican que el proceso se encontrará en el estado 1 el 2.49% de las veces, en el estado 2 el 93.74% de las veces, etc. El vector estacionario es el elemento central para los procesos irreducibles, y encontrarlo de manera eficiente es de crucial importancia. En [13] se destaca que en general uno debe evitar resolver estos sistemas de ecuaciones (3.26) y (3.48) de manera directa, ya que al aplicar métodos directos se pierde la estructura dispersa de la matriz y se pueden generar errores numéricos en el proceso. Se recomienda aplicar métodos iterativos para resolver el sistema de ecuaciones  $\mathbf{Ax}' = \mathbf{b}'$  como por ejemplo el método de Gauss-Siedel ó el método SOR, estos también tienen la ventaja de que se aplican

para problemas donde el espacio de estados es muy grande. Por el momento solo se aplicarán los métodos directos. A continuación se discutirán los procesos reducibles que tienen similitud con las cadenas de Markov finitas.

### 3.2.2 Procesos Reducibles

Asumiremos por el momento que el proceso tiene al menos un estado transitorio, la matriz de tasas de transición lo podremos reescribir en su forma reducible

$$\tilde{\mathbf{T}} = \begin{bmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix}$$

donde la submatriz  $\mathbf{C}$  contiene las tasas de transición entre estados recurrentes, la submatriz  $\mathbf{R}$  contiene las tasas de transición de los estados transitorios a los estados recurrentes y por último la submatriz  $\mathbf{Q}$  contiene las tasas de transición entre estados transitorios. Por las propiedades de los estados transitorios, se sabe que el proceso dejará de visitarlos en algún momento, por lo que se tiene que  $\lim_{t \rightarrow \infty} P_{i,j}(t) = 0$  para  $i, j \in T$ . En términos matriciales esta relación se traduce en

$$\lim_{t \rightarrow \infty} e^{\mathbf{Q}t} = \mathbf{0} \quad (3.50)$$

y esto se debe a que la matriz  $\mathbf{Q}$  es diagonalmente dominante y es el equivalente matricial a la expresión  $\lim_{t \rightarrow \infty} e^{-\beta t} = 0$  con  $\beta > 0$ . Se sabe que en algún momento el proceso saldrá de los estados transitorios y uno quiere estimar el tiempo esperado para que el proceso salga de los estados transitorios. Para ello, como en el caso discreto, se define la siguiente variable aleatoria

$$Y_{i,j}(k) = \begin{cases} 1 & \text{si } X(k) = j \text{ dado que } X(0) = i \\ 0 & \text{e.o.c} \end{cases}$$

para estados transitorios  $i, j \in T$ . Se define al tiempo de estancia del proceso por la variable aleatoria

$$N_{i,j}(t) = \int_0^t Y_{i,j}(\tau) d\tau$$

entonces lo que se desea es encontrar la esperanza de esta nueva variable aleatoria. Denotando a  $\mu_{i,j}(t) = E[N_{i,j}(t)]$  se obtiene lo siguiente

$$\begin{aligned} \mu_{i,j}(t) &= E[N_{i,j}(t)] = E \left[ \int_0^t Y_{i,j}(\tau) d\tau \right] \\ &= \int_0^t E[Y_{i,j}(\tau)] d\tau \\ &= \int_0^t P[X(\tau) = j \mid X(0) = i] d\tau \\ &= \int_0^t P_{i,j}(\tau) d\tau. \end{aligned}$$

Para sacar la integral del operador esperanza se utiliza el hecho de que es un operador lineal, aunque se necesitan más elementos matemáticos para dar una demostración formal. En términos matriciales tenemos que

$$\mathbf{M}(t) = \int_0^t e^{\mathbf{Q}\tau} d\tau. \quad (3.51)$$

Esta integral se puede resolver fácilmente usando el resultado en (2.14):

$$\begin{aligned} \mathbf{M}(t) &= \mathbf{Q}^{-1} e^{\mathbf{Q}\tau} \Big|_0^t \\ &= \mathbf{Q}^{-1} e^{\mathbf{Q}t} - \mathbf{Q}^{-1} e^{\mathbf{Q}(0)} \\ &= \mathbf{Q}^{-1} e^{\mathbf{Q}t} - \mathbf{Q}^{-1} \\ &= \mathbf{Q}^{-1} (e^{\mathbf{Q}t} - \mathbf{I}) \end{aligned}$$

procediendo como en el caso discreto, si definimos la matriz  $\mathbf{M} = \lim_{t \rightarrow \infty} \mathbf{M}(t)$  como el tiempo promedio que el proceso se encuentra en los estados transitorios antes de entrar a una clase recurrente. Recordando la condición (3.50) entonces se obtiene la siguiente matriz fundamental para procesos continuos

$$\mathbf{M} = -\mathbf{Q}^{-1} \quad (3.52)$$

Teniendo en cuenta que se queremos saber cuanto tiempo un estado transitorio  $i \in T$  estará en la clase transitoria solamente tenemos que sumar por renglones a la matriz  $\mathbf{M}$ , y a este vector lo denotaremos de nueva cuenta por  $\mathbf{M}^* = \mathbf{M}\mathbf{e}'$ . Recordando la matriz de tasas (3.41) teníamos que el estado 4 era un estado recurrente absorbente y los demás eran estados transitorios, tomando la permutación  $\sigma = (4 \ 1 \ 2 \ 3 \ 5)$  entonces obtenemos la matriz en forma reducible

$$\tilde{\mathbf{T}} = \mathbf{S}^T \mathbf{T} \mathbf{S} = \left[ \begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ \hline 0 & -2 & 2 & 0 & 0 \\ 0 & 3 & -9 & 2 & 4 \\ 0 & 0 & 1 & -1 & 0 \\ 1 & 0 & 0 & 3 & -4 \end{array} \right] \quad \mathbf{Q} = \begin{bmatrix} -2 & 2 & 0 & 0 \\ 3 & -9 & 2 & 4 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 3 & -4 \end{bmatrix}$$

con lo que se obtiene la matriz fundamental y el tiempo esperado que el proceso estuvo en los estados transitorios antes de entrar a una clase recurrente

$$\mathbf{M} = \begin{bmatrix} 2 & 1 & 5 & 1 \\ 1.5 & 1 & 5 & 1 \\ 1.5 & 1 & 6 & 1 \\ 1.125 & 0.75 & 4.5 & 1 \end{bmatrix} \quad \mathbf{M}^* = \begin{bmatrix} 9 \\ 8.5 \\ 9.5 \\ 7.375 \end{bmatrix}.$$

Estos resultados indican que el proceso estuvo en promedio en el estado 1 transitorio 9 unidades aproximadamente, estuvo en promedio 8.5 unidades de tiempo en el estado 2, y así

sucesivamente. Con esto tenemos una aproximación de cuanto tiempo debemos esperar para que el proceso salga de los estados transitorios y sea absorbido por alguna clase recurrente. Si se tienen varios estados recurrentes, la pregunta válida es cuál es el más probable que visite el proceso cuando salga de los estados transitorios, y para ello los estados recurrentes se convierten en estados absorbentes. Con el siguiente resultado se verá como obtener de manera fácil estas probabilidades.

**Teorema 3.34** *Supongamos que la matriz de tasas  $\mathbf{T}$  está en forma reducible de la siguiente manera*

$$\tilde{\mathbf{T}} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix},$$

La matriz está particionada con respecto al espacio de estados  $\mathbb{E} = \{\text{recurrentes, transitorios}\}$  y además los estados recurrentes son absorbentes. Por lo tanto la matriz de probabilidades de transición está dada por

$$\mathbf{P}(t) = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{Q}^{-1}(e^{\mathbf{Q}t} - \mathbf{I})\mathbf{R} & e^{\mathbf{Q}t} \end{bmatrix} \quad (3.53)$$

**Demostración.** Para demostrar este resultado lo que haremos es trabajar con las expansiones en series de Taylor de la función exponencial y con las potencias de la matriz de tasas de intensidades

$$\mathbf{P}(t) = \sum_{k=0}^{\infty} \frac{\mathbf{T}^k t^k}{k!} = \mathbf{I} + \mathbf{T}t + \frac{\mathbf{T}^2 t^2}{2!} + \frac{\mathbf{T}^3 t^3}{3!} + \dots$$

lo que debemos analizar ahora es la forma matricial que tienen las matrices  $\mathbf{T}^k$  para  $k \in \mathbb{N}$ . Las primeras potencias de esta matriz  $\tilde{\mathbf{T}}$  están dadas por:

$$\begin{aligned} \mathbf{T}^2 &= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{QR} & \mathbf{Q}^2 \end{bmatrix} \\ \mathbf{T}^3 &= \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{QR} & \mathbf{Q}^2 \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{Q}^2 \mathbf{R} & \mathbf{Q}^3 \end{bmatrix} \end{aligned}$$

y se puede demostrar por inducción que la potencia  $\mathbf{T}^k$  está dada por

$$\mathbf{T}^k = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{Q}^{k-1} \mathbf{R} & \mathbf{Q}^k \end{bmatrix}$$

sustituyendo estos valores en la expansión de  $\mathbf{P}(t)$  obtenemos

$$\begin{aligned} \mathbf{P}(t) &= \mathbf{I} + \mathbf{T}t + \frac{\mathbf{T}^2 t^2}{2!} + \frac{\mathbf{T}^3 t^3}{3!} + \dots \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{R} & \mathbf{Q} \end{bmatrix} t + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{QR} & \mathbf{Q}^2 \end{bmatrix} \frac{t^2}{2!} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{Q}^2 \mathbf{R} & \mathbf{Q}^3 \end{bmatrix} \frac{t^3}{3!} + \dots \\ &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{H}^* & e^{\mathbf{Q}t} \end{bmatrix} \end{aligned}$$

donde la matriz  $\mathbf{H}^*$  tiene la siguiente estructura

$$\mathbf{H}^* = \mathbf{R}t + \frac{\mathbf{Q}\mathbf{R}t^2}{2!} + \frac{\mathbf{Q}^2\mathbf{R}t^3}{3!} + \dots = \left( \sum_{k=1}^{\infty} \frac{\mathbf{Q}^{k-1}t^k}{k!} \right) \mathbf{R}$$

Entonces, si somos observadores, la expresión que está adentro de los paréntesis se parece mucho a la expansión en series de Taylor de la función exponencial, pero está desfasado un índice de la matriz  $\mathbf{Q}$ . Lo que haremos es multiplicar todo la igualdad por la matriz de intensidades  $\mathbf{Q}$  para actualizar los índices de la suma.

$$\mathbf{Q}\mathbf{H}^* = \mathbf{Q}\mathbf{R}t + \frac{\mathbf{Q}^2\mathbf{R}t^2}{2!} + \frac{\mathbf{Q}^3\mathbf{R}t^3}{3!} + \dots = \left( \sum_{k=1}^{\infty} \frac{\mathbf{Q}^k t^k}{k!} \right) \mathbf{R},$$

como la función exponencial empieza desde  $k = 0$  lo que nos falta por agregar es sumar y restar una matriz identidad de la misma dimensión que la matriz  $\mathbf{Q}$

$$\mathbf{Q}\mathbf{H}^* = \left( -\mathbf{I} + \mathbf{I} + \sum_{k=1}^{\infty} \frac{\mathbf{Q}^k t^k}{k!} \right) \mathbf{R} = \left( -\mathbf{I} + \sum_{k=0}^{\infty} \frac{\mathbf{Q}^k t^k}{k!} \right) \mathbf{R} = (-\mathbf{I} + e^{\mathbf{Q}t}) \mathbf{R}$$

y por último multiplicamos por la inversa de  $\mathbf{Q}$  por el lado izquierdo (la inversa existe ya que se trata de una matriz diagonal dominante)

$$\mathbf{H}^* = \mathbf{Q}^{-1} (-\mathbf{I} + e^{\mathbf{Q}t}) \mathbf{R}$$

con lo que se obtiene el resultado deseado. ■

A continuación mostraremos un razonamiento intuitivo para encontrar las probabilidades límite de absorción para pasar de los estados transitorios a los recurrentes. Recordemos que si tenemos dos estados transitorios  $i, j \in T$  entonces la probabilidad de que a la larga regresen a ellos mismos es cero. Si recordamos la relación (3.50) esto implica que las probabilidades límites de absorción de pasar de los estados transitorios a los estados recurrentes están dadas por la matriz  $\mathbf{F} = \mathbf{M}\mathbf{R}$

$$\lim_{t \rightarrow \infty} \mathbf{P}(t) = \left[ \begin{array}{cc|cc} \mathbf{I} & \mathbf{0} & & \\ -\mathbf{Q}^{-1}\mathbf{R} & \mathbf{0} & & \end{array} \right] = \left[ \begin{array}{cc|cc} \mathbf{I} & \mathbf{0} & & \\ \mathbf{M}\mathbf{R} & \mathbf{0} & & \end{array} \right]. \quad (3.54)$$

Veamos un ejemplo donde la matriz de tasas  $\mathbf{T}$  este compuesta por dos estados recurrentes y tres estados transitorios dada por

$$\tilde{\mathbf{T}} = \left[ \begin{array}{cc|ccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 5 & 3 & -9 & 1 & 0 \\ 1 & 0 & 6 & -10 & 3 \\ 1 & 2 & 3 & 0 & -6 \end{array} \right] \quad \mathbf{R} = \begin{bmatrix} 5 & 3 \\ 1 & 0 \\ 1 & 2 \end{bmatrix} \quad \mathbf{Q} = \begin{bmatrix} -9 & 1 & 0 \\ 6 & -10 & 3 \\ 3 & 0 & -6 \end{bmatrix} \quad (3.55)$$

y con estas dos submatrices encontramos las probabilidades límite de absorción:

$$\mathbf{F} = \mathbf{MR} = \begin{bmatrix} 0.6242 & 0.3758 \\ 0.6182 & 0.3818 \\ 0.4788 & 0.5212 \end{bmatrix}.$$

Con estos resultados encontramos que si empezamos en el estado 3 transitorio, entonces la probabilidad a la larga de ser absorbido por el estado 1 es del 62.42% y de ser absorbido por el estado 2 es el 37.58%. Si el proceso inicia en el estado 4 transitorio entonces la probabilidad de ser absorbido por el estado 1 es del 61.82% y por el estado 2 es del 38.18%, y así respectivamente. Al igual que con las cadenas finitas, obtener estas dos submatrices es muy importante para un análisis del proceso, por lo que se muestra un código similar visto anteriormente.

```
function [clases,R,Q]=extraccionRQcmtc(T)

% ESTA FUNCION EXTRAER LAS MATRICES "R" Y "Q" DE UN PROCESO
% REDUCIBLE CON AL MENOS UN ESTADO TRANSITORIO CON MATRIZ
% DE TASAS "T". PARA HACER EL ANALISIS SE UTILIZARA
% LA FUNCION "analisis_estados.m". LA FUNCION REGRESA A LA
% MATRIZ DE TASAS "T" EN SU FORMA REDUCIBLE. EL VECTOR
% "clases" TIENE LA SIGUIENTE ESTRUCTURA
%
% clases{1} = estados recurrentes
% clases{2} = estados transitorios
%
% FORMA DE UTILIZARSE [clases,R,Q]=extraccionRQcmtc(P)

[n,m]=size(T);
if(n~=m)
    error(' LA MATRIZ DE TASAS NO ES CUADRADA!!! ')
end

% Se encuentra la matriz de saltos auxiliar (solo las
% entradas positivas se analizan)
J = matriz_salto_aux(T);

% Se analiza la matriz
[rec,trans,clas_rec]=analisis_estados(J);
Rn=length(rec); Tn=length(trans); Cn=length(clas_rec);
clases{1}=rec; clases{2} = trans;

% Se analiza los tipos de formas que puede
% tener la matriz
if( Cn==1 && Tn==0 )
    error(' El proceso es irreducible!!! ')
end
```

```

elseif(Cn>1 && Tn==0)
    disp(' El proceso tiene varias clases recurrentes')
    error(' y ningun estado transitorio!!! ')
end

% Encontramos la permutacion sigma = [Rec Tran] y la matriz
% en su forma reducible y la extraccion de matrices
sigma = [ rec trans ]; T = T(sigma,sigma);
ext=(Rn+1):n; R=T(ext,1:Rn); Q=T(ext,ext);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function J=matriz_salto_aux(T)

[n,m]=size(T);

J=zeros(n);
% Buscamos los estados absorbentes primero
absors=find(diag(T)==0); J(absors,absors)=1;

% Las demas entradas donde es positivo se cambiaran por unos
posit=find(T>0); J(posit)=1;

```

Para ver como funciona el código, utilicemos la última matriz de tasas  $\mathbf{T}$  analizada en (3.55)

```

>> Ts=[0 0 0 0 0; ...; 1 0 6 -10 3; 1 2 3 0 -6]
>> [clases,R,Q]=extraccionRQcmtc(Ts);
>> R, Q
R =
     5     3
     1     0
     1     2

Q =
    -9     1     0
     6    -10     3
     3     0    -6

```

Con este último código se concluye con el análisis básico de los procesos de Markov reducibles. Al concluir este tema se vio la importancia de la matriz de tasas en la estructura del proceso, ya que con esta matriz se obtiene toda la información fundamental del proceso, y guarda una similitud con la matriz de transición en una cadena de Markov discreta.

# Capítulo 4

## Algunos modelos y aplicaciones

En este capítulo se trabajarán algunos modelos básicos tanto de los procesos discretos como los de tiempo continuo.

### 4.1 Redes estocásticas

Una red estocástica consiste en poder juntar dos ó más procesos estocásticos en uno solo y que estos procesos trabajen de manera paralela, para después poder modificar (de ser necesario) el comportamiento de la matriz de transición global (ó matriz de tasas) y obtener información particular del proceso. En términos prácticos, si se tienen dos cadenas de Markov discretas  $\{X_n\}_{n \in \mathbb{N}}$  y  $\{Y_n\}_{n \in \mathbb{N}}$  entonces una red estocástica discreta consiste en construir el proceso  $\{(X_n, Y_n)\}_{n \in \mathbb{N}}$  que es un proceso que trabaja en paralelo. Lo mismo sucede si se tienen dos procesos de Markov  $\{X(t)\}_{t \in \mathbb{R}}$  y  $\{Y(t)\}_{t \in \mathbb{R}}$  entonces la red estaría conformada por el proceso  $\{(X(t), Y(t))\}_{t \in \mathbb{R}}$ . Para poder encontrar las probabilidades de transición (o la matriz de tasas) es necesario trabajar con dos operadores importantes, los cuales se conocen como el **producto/suma tensorial de Kronecker**<sup>1</sup>.

**Definición 4.1** Sean  $\mathbf{A} \in \mathbb{R}^{n \times m}$  y  $\mathbf{B} \in \mathbb{R}^{p \times q}$  matrices rectangulares, entonces el **producto tensorial de Kronecker**  $\mathbf{A} \otimes \mathbf{B} = \mathbf{P} \in \mathbb{R}^{np \times mq}$  está definida por:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1m}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2m}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & a_{nm}\mathbf{B} \end{bmatrix} \quad (4.1)$$

---

<sup>1</sup>Leopold Kronecker (1823-1891). Matemático alemán. Su trabajo más en la rama de la teoría de los números y en álgebra abstracta. Su frase más importante fue "Dios hizo los números naturales, el resto es obra del hombre". Es reconocido por varios resultados matemáticos como la delta de Kronecker, el producto de Kronecker, entre otros.

**Definición 4.2** Sean  $\mathbf{A} \in \mathbb{R}^{n \times n}$  y  $\mathbf{B} \in \mathbb{R}^{m \times m}$  matrices cuadradas, entonces la **suma tensorial de Kronecker**  $\mathbf{A} \oplus \mathbf{B} = \mathbf{S} \in \mathbb{R}^{nm \times nm}$  está definida por

$$\boxed{\mathbf{A} \oplus \mathbf{B} = \mathbf{A} \otimes \mathbf{I}_m + \mathbf{I}_n \otimes \mathbf{B}} \quad (4.2)$$

donde  $\mathbf{I}_k$  es la matriz identidad de tamaño  $k$ .

Para entender estos dos operadores se verán dos ejemplos sencillos. Definamos las siguientes matrices (en este caso ambas cuadradas)

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 10 & 9 & 8 \\ 7 & 6 & 5 \\ 4 & 3 & 2 \end{bmatrix}$$

entonces el producto de Kronecker está dado por:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} \mathbf{B} & 2\mathbf{B} \\ 3\mathbf{B} & 4\mathbf{B} \end{bmatrix} = \left[ \begin{array}{ccc|ccc} 10 & 9 & 8 & 20 & 18 & 16 \\ 7 & 6 & 5 & 14 & 12 & 10 \\ 4 & 3 & 2 & 8 & 6 & 4 \\ \hline 30 & 27 & 24 & 40 & 36 & 32 \\ 21 & 18 & 15 & 28 & 24 & 20 \\ 12 & 9 & 6 & 16 & 12 & 8 \end{array} \right]$$

y la suma de Kronecker está dada por:

$$\begin{aligned} \mathbf{A} \oplus \mathbf{B} &= \mathbf{A} \otimes \mathbf{I}_3 + \mathbf{I}_2 \otimes \mathbf{B} \\ &= \begin{bmatrix} 11 & 9 & 8 & 2 & 0 & 0 \\ 7 & 7 & 5 & 0 & 2 & 0 \\ 4 & 3 & 3 & 0 & 0 & 2 \\ 3 & 0 & 0 & 14 & 9 & 8 \\ 0 & 3 & 0 & 7 & 10 & 5 \\ 0 & 0 & 3 & 4 & 3 & 6 \end{bmatrix} \end{aligned}$$

Con esto se aprecia que los operadores de Kronecker aumentan considerablemente el tamaño de las matrices con las que se está trabajando. Algunas propiedades importantes de los operadores de Kronecker se muestran a continuación, una demostración formal se puede checar en Broxson [46], Horn [45] ó Van Loan [47]

**Proposición 4.3** Los operadores de Kronecker  $\otimes$  y  $\oplus$  cumplen las siguientes propiedades

(en la multiplicación de matrices deben tener las dimensiones adecuadas)

$$\begin{aligned}
 1) \quad & \mathbf{A} \otimes \mathbf{B} \neq \mathbf{B} \otimes \mathbf{A} \\
 2) \quad & (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) \\
 3) \quad & (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD}) \\
 4) \quad & (\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C} \\
 5) \quad & \mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C} \\
 6) \quad & (\lambda \mathbf{A}) \otimes \mathbf{B} = \mathbf{A} \otimes (\lambda \mathbf{B}) = \lambda(\mathbf{A} \otimes \mathbf{B})
 \end{aligned} \tag{4.3}$$

Estas son algunas propiedades importantes sobre el producto tensorial de Kronecker. Una propiedad fundamental es que la matriz exponencial de la suma tensorial es igual al producto tensorial de las matrices exponenciales, y se demuestra a continuación.

**Teorema 4.4** Sean  $\mathbf{A} \in \mathbb{R}^{n \times n}$  y  $\mathbf{B} \in \mathbb{R}^{m \times m}$  dos matrices cuadradas cualesquiera, entonces se cumple

$$e^{\mathbf{A} \oplus \mathbf{B}} = e^{\mathbf{A}} \otimes e^{\mathbf{B}} \tag{4.4}$$

*Demostración.* Para esto se demuestra primero que las matrices  $\mathbf{A} \otimes \mathbf{I}_m$  y  $\mathbf{I}_n \otimes \mathbf{B}$  conmutan, para ello se utiliza la propiedad 3) de (4.3)

$$\begin{aligned}
 (\mathbf{A} \otimes \mathbf{I}_m)(\mathbf{I}_n \otimes \mathbf{B}) &= (\mathbf{I}_n \otimes \mathbf{B})(\mathbf{A} \otimes \mathbf{I}_m) \\
 \implies (\mathbf{AI}_n) \otimes (\mathbf{I}_m \mathbf{B}) &= (\mathbf{I}_n \mathbf{A}) \otimes (\mathbf{BI}_m) \\
 \implies \mathbf{A} \otimes \mathbf{B} &= \mathbf{A} \otimes \mathbf{B}
 \end{aligned}$$

y esto implica que si conmutan, por lo tanto se obtiene que

$$\exp(\mathbf{A} \otimes \mathbf{I}_m + \mathbf{I}_n \otimes \mathbf{B}) = \exp(\mathbf{A} \otimes \mathbf{I}_m) \exp(\mathbf{I}_n \otimes \mathbf{B})$$

tomando la expansión en series de Taylor de cada una de ellas se obtiene

$$\exp(\mathbf{A} \otimes \mathbf{I}_m) \exp(\mathbf{I}_n \otimes \mathbf{B}) = \left( \sum_{k=0}^{\infty} \frac{(\mathbf{A} \otimes \mathbf{I}_m)^k}{k!} \right) \left( \sum_{q=0}^{\infty} \frac{(\mathbf{I}_n \otimes \mathbf{B})^q}{q!} \right)$$

y de nueva cuenta por la propiedad 3) de (4.3) se obtiene que  $(\mathbf{A} \otimes \mathbf{I}_m)^k = (\mathbf{A}^k \otimes \mathbf{I}_m)$  y  $(\mathbf{I}_n \otimes \mathbf{B})^q = (\mathbf{I}_n \otimes \mathbf{B}^q)$  por lo que se obtiene

$$\exp(\mathbf{A} \otimes \mathbf{I}_m) \exp(\mathbf{I}_n \otimes \mathbf{B}) = \left( \sum_{k=0}^{\infty} \frac{(\mathbf{A}^k \otimes \mathbf{I}_m)}{k!} \right) \left( \sum_{q=0}^{\infty} \frac{(\mathbf{I}_n \otimes \mathbf{B}^q)}{q!} \right)$$

desarrollando los primeros términos de cada serie se obtiene

$$\left( (\mathbf{I}_n \otimes \mathbf{I}_m) + (\mathbf{A} \otimes \mathbf{I}_m) + \frac{(\mathbf{A}^2 \otimes \mathbf{I}_m)}{2!} + \dots \right) \left( (\mathbf{I}_n \otimes \mathbf{I}_m) + (\mathbf{I}_n \otimes \mathbf{B}) + \frac{(\mathbf{I}_n \otimes \mathbf{B}^2)}{2!} + \dots \right)$$

haciendo esta multiplicación matricial se obtiene lo siguiente

$$\begin{aligned} &= (\mathbf{I}_n^2 \otimes \mathbf{I}_m^2) + (\mathbf{I}_n \otimes \mathbf{I}_m) (\mathbf{I}_n \otimes \mathbf{B}) + (\mathbf{I}_n \otimes \mathbf{I}_m) \frac{(\mathbf{I}_n \otimes \mathbf{B}^2)}{2!} + \dots \\ &+ (\mathbf{A} \otimes \mathbf{I}_m) (\mathbf{I}_n \otimes \mathbf{I}_m) + (\mathbf{A} \otimes \mathbf{I}_m) (\mathbf{I}_n \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{I}_m) \frac{(\mathbf{I}_n \otimes \mathbf{B}^2)}{2!} + \dots \\ &\vdots \end{aligned}$$

y se vuelve a utilizar la propiedad 3) mencionada anteriormente junto con las propiedades 4), 5) y 6)

$$\begin{aligned} &= (\mathbf{I}_n \otimes \mathbf{I}_m) + (\mathbf{I}_n \otimes \mathbf{B}) + \left( \mathbf{I}_n \otimes \frac{\mathbf{B}^2}{2!} \right) + \dots \\ &+ (\mathbf{A} \otimes \mathbf{I}_m) + (\mathbf{A} \otimes \mathbf{B}) + \left( \mathbf{A} \otimes \frac{\mathbf{B}^2}{2!} \right) + \dots \\ &\vdots \\ &= \mathbf{I}_n \otimes \left( \mathbf{I}_m + \mathbf{B} + \frac{\mathbf{B}^2}{2!} + \dots \right) + \mathbf{A} \otimes \left( \mathbf{I}_m + \mathbf{B} + \frac{\mathbf{B}^2}{2!} + \dots \right) + \dots \\ &= \mathbf{I}_n \otimes e^{\mathbf{B}} + \mathbf{A} \otimes e^{\mathbf{B}} + \frac{\mathbf{A}^2}{2!} \otimes e^{\mathbf{B}} + \dots \\ &= \left( \mathbf{I}_n + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \dots \right) \otimes e^{\mathbf{B}} \\ &= e^{\mathbf{A}} \otimes e^{\mathbf{B}} \end{aligned}$$

que es el resultado que se desea obtener. ■

Estos elementos nos permitirán encontrar las matrices y tasas de procesos paralelos.

**Teorema 4.5** Sean  $\{X_k\}_{k \in \mathbb{N}}$  y  $\{Y_k\}_{k \in \mathbb{N}}$  dos cadenas de Markov finitas e independientes con espacios de estados  $\mathbb{E}_1$  y  $\mathbb{E}_2$  y matrices de transición  $\mathbf{P} \in \mathbb{R}^{n \times n}$  y  $\mathbf{Q} \in \mathbb{R}^{m \times m}$  respectivamente. Entonces el proceso  $\{\mathbf{Z}_k\}_{k \in \mathbb{N}}$  definido como

$$\{\mathbf{Z}_k\}_{k \in \mathbb{N}} = \{(X_k, Y_k)\}_{k \in \mathbb{N}}$$

es una cadena de Markov con espacio de estados  $\mathbb{E}_1 \times \mathbb{E}_2$  y matriz de transición  $\mathbf{Z} \in \mathbb{R}^{nm \times nm}$  dada por:

$$\boxed{\mathbf{Z} = \mathbf{P} \otimes \mathbf{Q}} \quad (4.5)$$

**Demostración.** Como son independientes ambos procesos se obtiene que

$$Z_{(i_0, i), (j_0, j)} = P_{i_0, j_0} Q_{i, j}$$

Fijemos dos estados cualesquiera  $i_0, j_0$  (pueden ser los mismos) para los estados de la primer cadena  $\{X_k\}_{k \in \mathbb{N}}$  y calculemos  $P_{i_0, j_0} Q_{i, j}$  para  $i, j = 1, \dots, m$ . La estructura por bloque para cada par  $i_0, j_0$  estaría dada de la siguiente manera

$$\begin{bmatrix} P_{i_0, j_0} Q_{1,1} & P_{i_0, j_0} Q_{1,2} & \cdots & P_{i_0, j_0} Q_{1,m} \\ P_{i_0, j_0} Q_{2,1} & P_{i_0, j_0} Q_{2,2} & \cdots & P_{i_0, j_0} Q_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ P_{i_0, j_0} Q_{m,1} & P_{i_0, j_0} Q_{m,2} & \cdots & P_{i_0, j_0} Q_{m,m} \end{bmatrix}$$

este es el  $(i_0, j_0)$  bloque de Kronecker. Entonces poniendolo en su forma general

$$\mathbf{Z} = \begin{bmatrix} P_{1,1} \mathbf{Q} & P_{1,2} \mathbf{Q} & \cdots & P_{1,n} \mathbf{Q} \\ P_{2,1} \mathbf{Q} & P_{2,2} \mathbf{Q} & \cdots & P_{2,n} \mathbf{Q} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n,1} \mathbf{Q} & P_{n,2} \mathbf{Q} & \cdots & P_{n,n} \mathbf{Q} \end{bmatrix} = \mathbf{P} \otimes \mathbf{Q}$$

Falta demostrar que la matriz  $\mathbf{Z}$  es una matriz estocástica. Tomando un "renglón" cualquiera de la matriz  $\mathbf{Z}$  y multiplicado por el vector columna de 1's obtenemos

$$\begin{aligned} \hat{\mathbf{Z}}_{i,*} \mathbf{e}' &= (P_{i,1} \mathbf{Q} + P_{i,2} \mathbf{Q} + \dots + P_{i,n} \mathbf{Q}) \mathbf{e}' \\ &= P_{i,1} \mathbf{Q} \mathbf{e}' + P_{i,2} \mathbf{Q} \mathbf{e}' + \dots + P_{i,n} \mathbf{Q} \mathbf{e}' \\ &= P_{i,1} \mathbf{e}' + P_{i,2} \mathbf{e}' + \dots + P_{i,n} \mathbf{e}' \\ &= \mathbf{P}_{i,*} \mathbf{e}' = 1 \end{aligned}$$

lo que implica que la matriz  $\mathbf{Z}$  es estocástica, y representa la matriz de transición del proceso  $\{(X_k, Y_k)\}_{k \in \mathbb{N}}$ . Es importante notar que los estados se ponen de forma lexicográfica. ■

**Teorema 4.6** Sean  $\{X(t)\}_{t \in \mathbb{R}^+}$  y  $\{Y(t)\}_{t \in \mathbb{R}^+}$  dos procesos de Markov finitos e independientes con espacios de estados  $\mathbb{E}_1$  y  $\mathbb{E}_2$  y matrices de tasas  $\mathbf{T}_1 \in \mathbb{R}^{n \times n}$  y  $\mathbf{T}_2 \in \mathbb{R}^{m \times m}$  respectivamente. Entonces el proceso  $\{\mathbf{Z}(t)\}_{t \in \mathbb{R}^+}$  definido como

$$\{\mathbf{Z}(t)\}_{t \in \mathbb{R}^+} = \{(X(t), Y(t))\}_{t \in \mathbb{R}}$$

es un proceso de Markov con espacio de estados  $\mathbb{E}_1 \times \mathbb{E}_2$  y cuya matriz de tasas  $\mathbf{T} \in \mathbb{R}^{nm \times nm}$  está dada por:

$$\mathbf{T} = \mathbf{T}_1 \otimes \mathbf{I}_m + \mathbf{I}_n \otimes \mathbf{T}_2 \quad (4.6)$$

**Demostración.** Se sabe que el proceso  $\{X(t)\}_{t \in \mathbb{R}^+}$  tiene como matriz de transición a  $\mathbf{P}(t) = e^{\mathbf{T}_1 t}$  y que el proceso  $\{Y(t)\}_{t \in \mathbb{R}^+}$  tiene matriz de transición  $\mathbf{Q}(t) = e^{\mathbf{T}_2 t}$ , entonces por el teorema anterior el proceso  $\{\mathbf{Z}(t)\}_{t \in \mathbb{R}^+}$  tiene matriz de transición

$$\begin{aligned} \mathbf{R}(t) &= \mathbf{P}(t) \otimes \mathbf{Q}(t) \\ &= e^{\mathbf{T}_1 t} \otimes e^{\mathbf{T}_2 t} \end{aligned}$$

Por último, usando el resultado (4.4) se tiene que

$$e^{\mathbf{T}_1 t} \otimes e^{\mathbf{T}_2 t} = e^{(\mathbf{T}_1 \oplus \mathbf{T}_2) t}$$

y con esto se concluye que la matriz de tasas del proceso conjunto es  $\mathbf{T}_1 \otimes \mathbf{I}_2 + \mathbf{I}_1 \otimes \mathbf{T}_2$ , que es lo que se quería demostrar. ■

Estos resultados son importantes ya que permiten encontrar de manera práctica las matrices de transición y de tasas de procesos paralelos. A continuación se muestran sin demostración resultados similares para cuando se trabaja con más procesos simultáneamente.

**Teorema 4.7** Sean  $\{X_n^{(1)}\}_{n \in \mathbb{N}}, \dots, \{X_n^{(k)}\}_{n \in \mathbb{N}}$   $k$  cadenas de Markov finitas e independientes con espacio de estados  $\mathbb{E}_1, \mathbb{E}_2, \dots, \mathbb{E}_k$  y matrices de transición  $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_k$  respectivamente. Entonces el proceso definido por

$$\{\mathbf{Z}_n\}_{n \in \mathbb{N}} = \{(X_n^{(1)}, \dots, X_n^{(k)})\}_{n \in \mathbb{N}}$$

es una cadena de Markov con espacio de estados  $\mathbb{E}_1 \times \mathbb{E}_2 \times \dots \times \mathbb{E}_k$  y matriz de transición dada por:

$$\mathbf{Z} = \mathbf{P}_1 \otimes \mathbf{P}_2 \otimes \dots \otimes \mathbf{P}_k \tag{4.7}$$

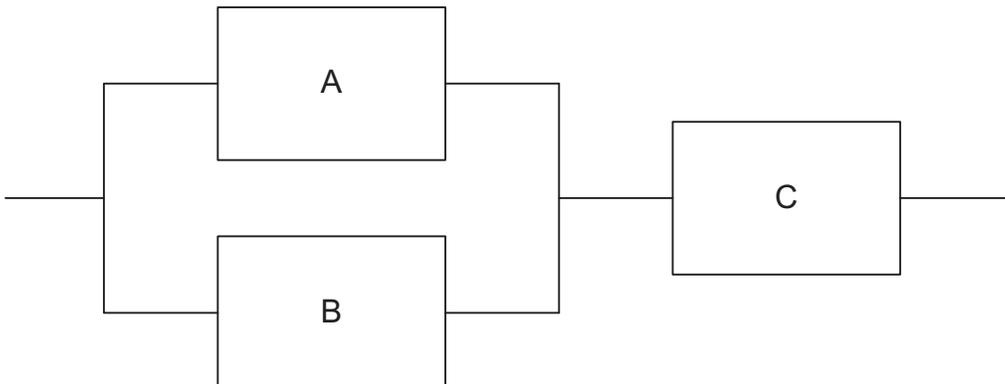
**Teorema 4.8** Sean  $\{X_1(t)\}_{t \in \mathbb{R}^+}, \dots, \{X_k(t)\}_{t \in \mathbb{R}^+}$   $k$  procesos de Markov finitos e independientes con espacio de estados  $\mathbb{E}_1, \mathbb{E}_2, \dots, \mathbb{E}_k$  y con matrices de tasas  $\mathbf{T}_1 \in \mathbb{R}^{n_1 \times n_1}, \mathbf{T}_2 \in \mathbb{R}^{n_2 \times n_2}, \dots, \mathbf{T}_k \in \mathbb{R}^{n_k \times n_k}$  respectivamente. Entonces el proceso definido por

$$\{\mathbf{Z}(t)\}_{t \in \mathbb{R}^+} = \{(X_1(t), \dots, X_k(t))\}_{t \in \mathbb{R}^+}$$

es un proceso de Markov con espacio de estados  $\mathbb{E}_1 \times \mathbb{E}_2 \times \dots \times \mathbb{E}_k$  y matriz de tasas dada por:

$$\mathbf{T} = \bigoplus_{i=1}^k \mathbf{T}_i = \sum_{i=1}^k \mathbf{I}_{n_1} \otimes \dots \otimes \mathbf{I}_{n_{i-1}} \otimes \mathbf{T}_i \otimes \mathbf{I}_{n_{i+1}} \otimes \dots \otimes \mathbf{I}_{n_k} \tag{4.8}$$

Para entender el significado práctico de la matriz de tasas para  $k$  procesos en paralelo, analizaremos en primera instancia un modelo básico para encontrar el tiempo aproximado de vida de un circuito eléctrico con la siguiente configuración:



donde cada bloque representa un circuito eléctrico, y tiene una configuración en paralelo para el paso de corriente en  $A$  y  $B$ , por lo que si no fluye corriente en alguno de ellos sigue funcionando esta parte, pero está en una configuración en serie con  $C$ , por lo que si deja de funcionar el bloque  $C$  entonces todo el proceso deja de funcionar. El funcionamiento de cada componente se puede modelar de manera independiente como un proceso con dos estados  $\mathbb{E} = \{0, 1\}$  donde el estado 0 representa que el componente no funciona y el estado 1 representa que funciona. Aquí lo importante es que si se asume que el tiempo de vida de cada componente se puede modelar como una distribución exponencial (1.26), entonces se puede modelar como un proceso de Markov. En estos casos mediante experimentación se pueden obtener estimaciones del tiempo de vida de cada componente. Por ejemplo, si tenemos la información de que un componente tiene una vida promedio de 40 horas, entonces la tasa de la distribución exponencial es  $\lambda = 1/40 = 0.025 \text{ hr}^{-1}$  y se le conoce como la **tasa de fallo**. En nuestro caso en particular se puede pensar que el componente  $C$  debe tener un tiempo de vida promedio mayor que los componentes  $A$  y  $B$ , por lo que su tasa de fallo es mucho menor. También se puede suponer que si algún componente falla, entonces el tiempo promedio de reparación es igual para los 3 componentes. Como valores en particular se asumirá que el tiempo promedio de reemplazo es de 2 días, para los componentes  $A$  y  $B$  el tiempo promedio de vida es de 200 días y para el componente  $C$  el tiempo promedio de vida es de 500 días, por lo que las matrices de tasas de intensidad de cada componente estarían dados por:

$$\begin{aligned} \mathbf{T}_A &= \mathbf{T}_B = \begin{bmatrix} -1/2 & 1/2 \\ 1/200 & -1/200 \end{bmatrix} = \begin{bmatrix} -0.5 & 0.5 \\ 0.005 & -0.005 \end{bmatrix} \\ \mathbf{T}_C &= \begin{bmatrix} -1/2 & 1/2 \\ 1/500 & -1/500 \end{bmatrix} = \begin{bmatrix} -0.5 & 0.5 \\ 0.002 & -0.002 \end{bmatrix} \end{aligned}$$

El proceso estocástico  $\{\mathbf{Z}_{PS}(t)\}_{t \in \mathbb{R}^+} = \{(A(t), B(t), C(t))\}_{t \in \mathbb{R}^+}$  representa la configuración de todo el sistema en el momento  $t$  y el espacio de estados estaría dado por

$$\begin{aligned} \mathbb{E}_{PS} &= \{(0, 0, 0), (0, 0, 1), (0, 1, 0), (0, 1, 1), (1, 0, 0), (1, 0, 1), (1, 1, 0), (1, 1, 1)\} \\ \mathbb{E}_{PS}^* &= \{1^*, 2^*, 3^*, 4^*, 5^*, 6^*, 7^*, 8^*\} \end{aligned}$$

el cambio de nombre de los estados es para facilitar el análisis en Mat/Oct. Asumiendo que no hay una configuración especial la matriz de tasas estaría dada por

$$\begin{aligned} \mathbf{T}_{PS} &= \mathbf{T}_A \otimes \mathbf{I}_2 \otimes \mathbf{I}_2 + \mathbf{I}_2 \otimes \mathbf{T}_B \otimes \mathbf{I}_2 + \mathbf{I}_2 \otimes \mathbf{I}_2 \otimes \mathbf{T}_C \\ &= \begin{bmatrix} -1.5 & 0.5 & 0.5 & 0 & 0.5 & 0 & 0 & 0 \\ 0.002 & -1.002 & 0 & 0.5 & 0 & 0.5 & 0 & 0 \\ 0.005 & 0 & -1.005 & 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0.005 & 0.002 & -0.5070 & 0 & 0 & 0 & 0.5 \\ 0.005 & 0 & 0 & 0 & -1.005 & 0.5 & 0.5 & 0 \\ 0 & 0.005 & 0 & 0 & 0.002 & -0.5070 & 0 & 0.5 \\ 0 & 0 & 0.005 & 0 & 0.005 & 0 & -0.51 & 0.5 \\ 0 & 0 & 0 & 0.005 & 0 & 0.005 & 0.002 & -0.012 \end{bmatrix} \end{aligned}$$

En la construcción de esta matriz global de tasas se asume que todo el proceso está en paralelo, no hay ninguna configuración especial. En nuestro caso, al querer que el circuito

se encuentre en un proceso combinado paralelo y en serie, se deben transformar los estados  $1^*$ ,  $2^*$ ,  $3^*$ ,  $5^*$  y  $7^*$  en absorbentes, para analizar el tiempo medio de confiabilidad del sistema y con cuál es la configuración más probable con la que el sistema deje de funcionar. Todo este análisis se puede resolver fácilmente con Mat/Oct.

```
>> TA=[-1/2 1/2; 1/200 -1/200]; TB=TA; TC=[-1/2 1/2; 1/500 -1/500];
>> I2 = eye(2);
>> Tps=kron(TA,kron(I2,I2)) + kron(I2,kron(TB,I2)) + kron(I2,kron(I2,TC));
>> absorbs=[1 2 3 5 7];
>> for k=absorbs
Tps(k,:)=zeros(1,8);
end
>> [clases,R,Q]=extraccionRQcmtc(Tps);

>> clases{1} % Los estados recurrentes
ans =
     1     2     3     5     7

>> clases{2} % Los estados transitorios
ans =
     4     6     8

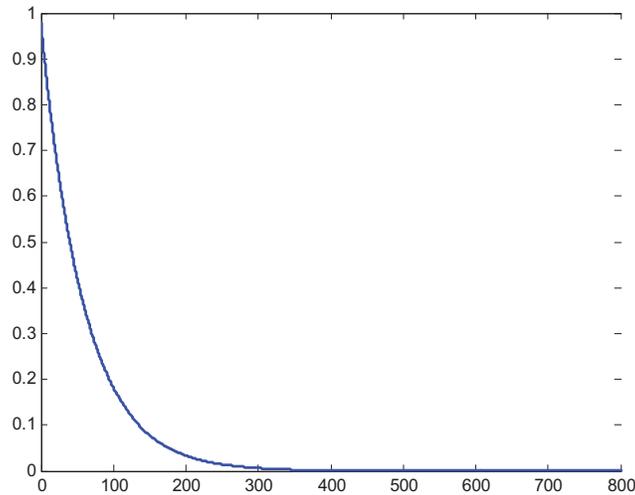
>> M=-inv(Q); Mast=sum(M,2)
Mast =
    472.3247
    472.3247
    476.9373

>> F=M*R
F =
     0    0.0554    0.0130    0.0091    0.9225
     0    0.0554    0.0091    0.0130    0.9225
     0    0.0461    0.0092    0.0092    0.9354
```

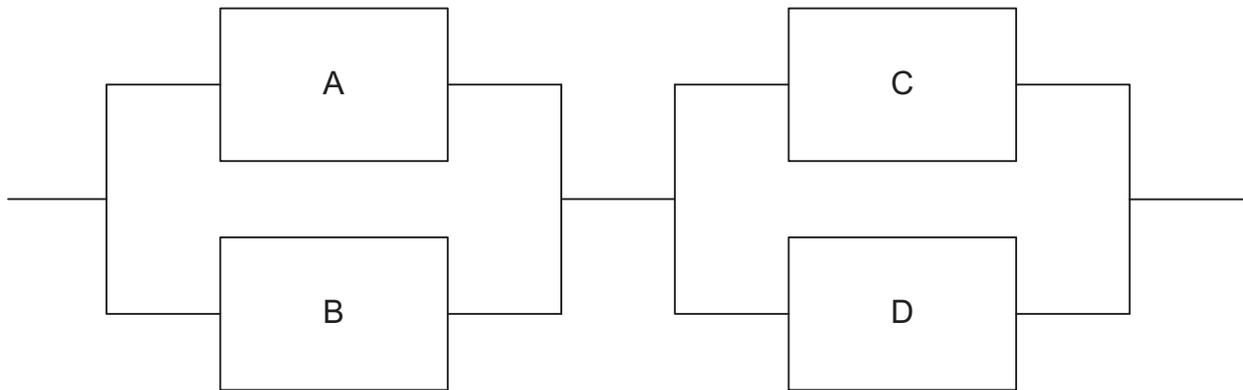
Nótese que el tiempo promedio de vida de todo el sistema es más del doble de el tiempo promedio de vida de los componentes  $A$  y  $B$ , que al construirlo en paralelo se espera de manera intuitiva que efectivamente se cumpla que aumente considerablemente el tiempo promedio de vida global del sistema. Las probabilidades de última absorción nos dicen en cuál configuración es más probable que el sistema se colapse

Estados	(0,0,0)	(0,0,1)	(0,1,0)	(1,0,0)	(1,1,0)
(0,1,1)	0	0.0554	0.0130	0.0091	0.9225
(1,0,1)	0	0.0554	0.0091	0.0130	0.9225
(1,1,1)	0	0.0461	0.0092	0.0092	0.9354

y estos resultados nos confirman de manera inicial que por las propiedades del sistema nunca fallarán los tres componentes simultáneamente, y que este se colapsará en el momento de que el componente  $C$  falle. Se puede modificar un poco el archivo m-file propuesto anteriormente *metodo\_Pade\_mejorado.m*, con la finalidad de que se pueda visualizar solo las probabilidades de transición  $P_{(1,1,1),(1,1,1)}(t)$ , es decir, solo las probabilidades de que el sistema se encuentre operando sin que falle ningún circuito. También se modificará inicialmente el tamaño de paso, el propuesto inicialmente es  $\Delta = 0.1$  y en este caso se tomará  $\Delta = 0.4$  y un intervalo de tiempo  $\tau = 800$



Esto nos muestra que el sistema no puede mantenerse siempre en óptimas condiciones, y esto se debe a las condiciones del tiempo de vida de los circuitos. Estas probabilidades se espera que puedan mejorar si se construye un sistema con 4 componentes, descritos en paralelo/paralelo, como se muestra en el siguiente diagrama.



Como esta el proceso en paralelo/paralelo, hay menos condiciones para que el sistema se colapse y deje de funcionar. Si tomamos ahora que las matrices de tasas para los componentes  $C$  y  $D$  tengan el mismo comportamiento que los componentes  $A$  y  $B$ , se obtiene ahora un proceso que tiene 16 estados, y cuya matriz de tasas global está dada por:

$$\mathbf{T}_{PP} = \mathbf{T}_A \otimes \mathbf{I}_2 \otimes \mathbf{I}_2 \otimes \mathbf{I}_2 + \mathbf{I}_2 \otimes \mathbf{T}_B \otimes \mathbf{I}_2 \otimes \mathbf{I}_2 + \mathbf{I}_2 \otimes \mathbf{I}_2 \otimes \mathbf{T}_C \otimes \mathbf{I}_2 + \mathbf{I}_2 \otimes \mathbf{I}_2 \otimes \mathbf{I}_2 \otimes \mathbf{T}_D$$

y el espacio de estados del proceso  $\{\mathbf{Z}_{PP}(t)\}_{t \in \mathbb{R}^+} = \{(A(t), B(t), C(t), D(t))\}_{t \in \mathbb{R}^+}$  está dado por:

$$\mathbb{E}_{PP} = \left\{ \begin{array}{l} (0, 0, 0, 0), (0, 0, 0, 1), (0, 0, 1, 0), (0, 0, 1, 1), (0, 1, 0, 0), (0, 1, 0, 1), (0, 1, 1, 0), (0, 1, 1, 1) \\ (1, 0, 0, 0), (1, 0, 0, 1), (1, 0, 1, 0), (1, 0, 1, 1), (1, 1, 0, 0), (1, 1, 0, 1), (1, 1, 1, 0), (1, 1, 1, 1) \end{array} \right\}$$

$$\mathbb{E}_{PP}^* = \{1^*, 2^*, 3^*, 4^*, 5^*, 6^*, 7^*, 8^*, 9^*, 10^*, 11^*, 12^*, 13^*, 14^*, 15^*, 16^*\}$$

Para esta configuración, los estados  $1^*, 2^*, 3^*, 4^*, 5^*, 9^*$  y  $13^*$  son los que causarían que el sistema se colapsará, por lo que en nuestro caso de nueva cuenta se harán absorbentes, y se analizará el tiempo promedio de vida de este nuevo sistema. Este análisis se hará de manera fácil en Mat/Oct

```
>> TA=[-1/2 1/2; 1/200 -1/200]; TB=TA; TC=TA; TD=TA;
>> I2=eye(2);
>> Tpp=kron(TA,kron(I2,kron(I2,I2))) + kron(I2,kron(TB,kron(I2,I2))) + ...
      kron(I2,kron(I2,kron(TC,I2))) + kron(I2,kron(I2,kron(I2,TD)));
>> absorbs=[1 2 3 4 5 9 13];
>> for k=absorbs
Tpp(k,:)=zeros(1,16);
end
>> [clases,R,Q]=extraccionRQcmtc(Tpp);

>> clases{1} % Los estados recurrentes
ans =
     1     2     3     4     5     9    13

>> clases{2} % Los estados transitorios
ans =
     6     7     8    10    11    12    14    15    16

>> M=-inv(Q); Mast=sum(M,2)
Mast =
    5051.45631067967
    5051.45631067967
    5100.97087378647
    5051.45631067967
    5051.45631067967
    5100.97087378647
    5100.97087378647
    5100.97087378647
    5100.97087378647
    5150.97087378647

>> F=M*R
F =
     0     0.0098     0.0048     0.4854     0.0098     0.0048     0.4854
     0     0.0048     0.0098     0.4854     0.0098     0.0048     0.4854
```

0	0.0049	0.0049	0.4951	0.0049	0.0048	0.4854
0	0.0098	0.0048	0.4854	0.0048	0.0098	0.4854
0	0.0048	0.0098	0.4854	0.0048	0.0098	0.4854
0	0.0049	0.0049	0.4951	0.0048	0.0049	0.4854
0	0.0049	0.0048	0.4854	0.0049	0.0049	0.4951
0	0.0048	0.0049	0.4854	0.0049	0.0049	0.4951
0	0.0049	0.0049	0.4903	0.0049	0.0049	0.4903

Con esto se observa un tiempo promedio de vida considerablemente mayor al que se tiene si solo se trabaja en paralelo/serie, los estados más probables para que el sistema colapse son el  $4^*$  y  $13^*$ , que en la configuración inicial son  $(0, 0, 1, 1)$  y  $(1, 1, 0, 0)$  respectivamente, y esto es que los componentes  $A$  y  $B$  ó los componentes  $C$  y  $D$  fallen simultáneamente. Con estos dos ejemplos se aprecia la capacidad de analizar sistemas más complejos, aunque una de las desventajas es que el espacio de estados crece de manera exponencial. Este modelo es importante para poder entender la interacción que pueden llegar a tener sistemas en paralelo, y entre más sistemas esten interactuando es más difícil poder encontrar la matriz global de tasas. Para ello se construye un archivo m-file que genere fácilmente la suma tensorial de Kronecker al introducir una cantidad arbitraria de matrices de tasas, ya que para ello solo es necesario utilizar la propiedad de que el producto tensorial de dos matrices identidades es igual a una matriz identidad cuya dimensión es el producto de las dimensiones menores.

$$\boxed{\mathbf{I}_n \otimes \mathbf{I}_m = \mathbf{I}_{nm}} \quad (4.9)$$

```
function Tglobal = generador_redes(varargin)

% ESTA FUNCION REGRESA LA MATRIZ GLOBAL DE TASAS "Tglobal"
% DADO QUE SE INTRODUJO UNA CANTIDAD FINITA DE MATRICES A1,A2,...
% QUE SE DEBE VERIFICAR QUE SEAN MATRICES CUADRADAS.
%
% FORMA DE UTILIZARSE: Tglobal = generador_redes(A1,A2,...)

% Para saber cuantas matrices se introdujeron
N = nargin;

% Construimos un vector auxiliar con las dimensiones de cada matriz
Tam = zeros(1,N);
for k=1:N
    Tam(k)=max(size(varargin{k}));
end

% Se analizan si se tienen dos matrices o mas
if( N==2 )
    Tglobal=kron(varargin{1},eye(Tam(2))) + ...
```

```

        ...kron(eye(Tam(1)),varargin{2});
else
    % Generamos inicialmente la suma global A1 k I2 k I3... k IN
    % y el ultimo I1 k I2 k I3 k ... k AN

    Tglobal = kron(varargin{1},speye(prod(Tam(2:end)))) + ...
        ... kron(speye(prod(Tam(1:N-1))),varargin{N});

    % Y en esta ultima parte se generan las matrices intermedias
    % I1 k I2 k ... k Aj k ... k IN

    for k=2:N-1
        Tglobal = Tglobal + kron(speye(prod(Tam(1:k-1))), ...
            ...kron(varargin{k},speye(prod(Tam(k+1:N)))));
    end
end

% La matriz Tglobal es sparse, por lo que se regresa su version "completa"
Tglobal=full(Tglobal);

```

Este archivo será de mucha utilidad debido a que se puede introducir cualquier cantidad de matrices de tasas y encontrar la matriz de tasas global, ya que si se observa el ejemplo de los cuatro circuitos era complicado estar introduciendo tantos operadores de Kronecker encajados que dificultaría el cálculo global usando la propiedad formal  $\mathbf{I}_{n_1} \otimes \mathbf{I}_{n_2} \otimes \cdots \otimes \mathbf{A}_k \otimes \cdots \otimes \mathbf{I}_{n_k}$ . Esto garantiza que se puede encontrar la matriz global de tasas de  $N$  procesos en paralelo de manera sencilla, aunque se debe tener en cuenta que entre más procesos se introduzcan y si el espacio de estados de cada proceso es muy grande guardar en memoria esta matriz global puede resultar computacionalmente costoso.

## 4.2 Teoría de Colas

La principal área de estudio de la teoría de colas es el análisis de los tiempos de espera en distintas situaciones donde se realicen "servicios". En estos casos existen distintas modalidades de como se realizan los servicios y los tiempos de su realización. Ejemplos básicos sobre este tipo de arribos se dan por ejemplo en la llegada de llamadas telefónicas a una central telefónica, cuando automovilistas solicitan servicio en una gasolinera, el arribo de clientes a la zona de cajas de un supermercado, entre otros. Cuando arriban clientes al sistema pueden hacerlo de manera de manera individual o por lotes. A su vez, cuando los clientes llegan al sistema, es necesario analizar el tiempo que se requiere para la realización del servicio, por lo que esto es el segundo elemento a considerar, ya que el tiempo de realización se puede considerar igual para cada cliente o puede variar de forma aleatoria. Un tercer elemento a considerar es el número de "estaciones" que se tienen para realizar el servicio, ya que se puede tener una sola estación de trabajo o varios a la vez, y en particular cada estación de servicio puede tener tiempos de realización diferentes (por ejemplo, en un super mercado una cajera novata se tardará más en atender a un cliente que una cajera con experiencia). Una

cuarta característica básica es el que se aplica para atender a los clientes, en un caso general se pueden aceptar "una infinidad" de clientes en el sistema, pero en la práctica se tiene una capacidad finita para la cola del sistema. Es común que a los clientes se les atiende como fueron arribando al sistema, lo que se conoce como "Primero Llega Primero Sale" (FIFO, "First In First Out" en inglés) y es en general la forma más común de atención. Hay otra forma de atender a los clientes, esto se debe a que algunos clientes tienen más prioridad que otros (por ejemplo, cuando se tiene un accidente automovilístico y se llega a una sala de emergencias, es atendido primero el accidentado antes que la gente que va solamente de consulta). A este tipo de servicio se le conoce como "Último Llega Primero Sale" (LIFO, "Last In First Out" en inglés). Como primera aproximación matemática a este tipo de problemas se plantean los procesos de nacimiento y muerte.

Un **proceso de nacimiento y muerte** se asume que se tiene una población de tamaño  $n$  y que solamente puede ocurrir un nacimiento ó una muerte en un intervalo corto de tiempo.

1. En una población de tamaño  $n$  al tiempo  $t$  la probabilidad de que ocurra un **nacimiento** en un intervalo de longitud  $h$  es  $\lambda_n h + o(h)$  y que no ocurra un nacimiento es de  $1 - \lambda_n h + o(h)$
2. En una población de tamaño  $n$  al tiempo  $t$  la probabilidad de que ocurra una **muerte** en un intervalo de longitud  $h$  es  $\mu_n h + o(h)$  y la probabilidad de que no ocurra una muerte es de  $1 - \mu_n h + o(h)$
3. El número de nacimientos es independiente al número de muertes para un mismo tamaño de población
4. Si al tiempo  $t$  el tamaño de la población es cero, entonces la probabilidad de que ocurra un deceso en el intervalo  $(t, t + h)$  es cero

La función  $o(\cdot)$  se puede entender como una función despreciable y cumple  $\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$ . Si se tiene un tamaño de población  $n$ , la tasa con la que nace una unidad es de  $\lambda_n$  y la tasa con la que muere una unidad es de  $\mu_n$ . Con ello se pueden obtener las siguientes probabilidades de transición

$$\begin{aligned} \text{a) } P_{n,n+1}(t, t+h) &= (\lambda_n h + o(h)) \cdot (1 - \mu_n h + o(h)) \\ &= \lambda_n h + o(h) \end{aligned}$$

$$\begin{aligned} \text{b) } P_{n,n-1}(t, t+h) &= (\mu_n h + o(h)) \cdot (1 - \lambda_n h + o(h)) \\ &= \mu_n h + o(h) \end{aligned}$$

$$\begin{aligned} \text{c) } P_{n,n}(t, t+h) &= (1 - \lambda_n h + o(h)) \cdot (1 - \mu_n h + o(h)) \\ &= 1 - (\lambda_n + \mu_n)h + o(h) \end{aligned}$$

Bajo estas hipótesis de modelación, se construyen las ecuaciones de Chapman-Kolmogorov hacia atrás y hacia adelante. En el primer caso se descompone el intervalo  $(0, t+h)$  en dos intervalos, los cuales son  $(0, h)$  y  $(h, t+h)$ , y se obtienen las **ecuaciones de Kolmogorov**

**hacia atrás.** Se asume que el proceso es homogéneo, es decir, que las probabilidades no dependen del intervalo que se este analizando. Por simplificación denotamos a las probabilidades  $P_{i,j}(t, t+h) = P_{i,j}(0, h) = P_{i,j}(h)$ . Utilizando las ecuaciones de Chapman-Kolmogorov

$$\begin{aligned}
 P_{i,j}(0, t+h) &= \sum_{k \in \mathbb{E}} P_{i,k}(0, h) P_{k,j}(h, t+h) \\
 &= \sum_{k \in \mathbb{E}} P_{i,k}(0, h) P_{k,j}(0, t) \\
 &= P_{i,i-1}(h) P_{i-1,j}(t) + P_{i,i}(h) P_{i,j}(t) + P_{i,i+1}(h) P_{i+1,j}(t) + o(h) \\
 &= \mu_i h P_{i-1,j}(t) + (1 - (\lambda_i + \mu_i)h) P_{i,j}(t) + \lambda_i h P_{i+1,j}(t) + o(h) \\
 &= \mu_i h P_{i-1,j}(t) + P_{i,j}(t) - (\lambda_i + \mu_i)h P_{i,j}(t) + \lambda_i h P_{i+1,j}(t) + o(h)
 \end{aligned}$$

después, pasando el término  $P_{i,j}(t)$  al lado izquierdo y dividiendo todo entre  $h$  se obtiene que

$$\frac{P_{i,j}(t+h) - P_{i,j}(t)}{h} = \mu_i P_{i-1,j}(t) - (\lambda_i + \mu_i) P_{i,j}(t) + \lambda_i P_{i+1,j}(t) + \frac{o(h)}{h}$$

y tomando por último el límite cuando  $h \rightarrow 0$  se obtiene las ecuaciones de Kolmogorov hacia atrás

$$\begin{aligned}
 &\begin{bmatrix} P'_{0,0}(t) & P'_{0,1}(t) & P'_{0,2}(t) & \cdots \\ P'_{1,0}(t) & P'_{1,1}(t) & P'_{1,2}(t) & \cdots \\ P'_{2,0}(t) & P'_{2,1}(t) & P'_{2,2}(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \dots \\
 \dots &= \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -(\mu_1 + \lambda_1) & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -(\mu_2 + \lambda_2) & \lambda_2 & 0 & \cdots \\ 0 & 0 & \mu_3 & -(\mu_3 + \lambda_3) & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} P_{0,0}(t) & P_{0,1}(t) & P_{0,2}(t) & \cdots \\ P_{1,0}(t) & P_{1,1}(t) & P_{1,2}(t) & \cdots \\ P_{2,0}(t) & P_{2,1}(t) & P_{2,2}(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}
 \end{aligned}$$

De manera similar se construyen las **ecuaciones de Kolmogorov hacia adelante**, tomando en cuenta ahora que el intervalo  $(0, t+h)$  se puede descomponer como dos intervalos disjuntos  $(0, t)$  y  $(t, t+h)$

$$\begin{aligned}
 P_{i,j}(0, t+h) &= \sum_{k \in \mathbb{E}} P_{i,k}(0, t) P_{k,j}(t, t+h) \\
 &= \sum_{k \in \mathbb{E}} P_{i,k}(0, t) P_{k,j}(0, h) \\
 &= P_{i,j-1}(t) P_{j-1,j}(h) + P_{i,j}(t) P_{j,j}(h) + P_{i,j+1}(t) P_{j+1,j}(h) + o(h) \\
 &= P_{i,j-1}(t) \lambda_{j-1} h + P_{i,j}(t) (1 - (\mu_j + \lambda_j)h) + P_{i,j+1}(t) \mu_{j+1} h + o(h) \\
 &= P_{i,j-1}(t) \lambda_{j-1} h + P_{i,j}(t) - (\mu_j + \lambda_j)h P_{i,j}(t) + P_{i,j+1}(t) \mu_{j+1} h + o(h)
 \end{aligned}$$

después, pasando el término  $P_{i,j}(t)$  al lado izquierdo y dividiendo todo entre  $h$  se obtiene que

$$\frac{P_{i,j}(t+h) - P_{i,j}(t)}{h} = P_{i,j-1}(t) \lambda_{j-1} - (\mu_j + \lambda_j) P_{i,j}(t) + P_{i,j+1}(t) \mu_{j+1} + \frac{o(h)}{h}$$

y por último tomando el límite cuando  $h \rightarrow 0$  se obtiene las ecuaciones de Kolmogorov hacia adelante

$$\begin{aligned} & \begin{bmatrix} P'_{0,0}(t) & P'_{0,1}(t) & P'_{0,2}(t) & \cdots \\ P'_{1,0}(t) & P'_{1,1}(t) & P'_{1,2}(t) & \cdots \\ P'_{2,0}(t) & P'_{2,1}(t) & P'_{2,2}(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} = \dots \\ \dots & = \begin{bmatrix} P_{0,0}(t) & P_{0,1}(t) & P_{0,2}(t) & \cdots \\ P_{1,0}(t) & P_{1,1}(t) & P_{1,2}(t) & \cdots \\ P_{2,0}(t) & P_{2,1}(t) & P_{2,2}(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -(\mu_1 + \lambda_1) & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -(\mu_2 + \lambda_2) & \lambda_2 & 0 & \cdots \\ 0 & 0 & \mu_3 & -(\mu_3 + \lambda_3) & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \end{aligned}$$

En ambos casos, la matriz de tasas  $\mathbf{T}$  del proceso de nacimiento y muerte está dada por

$$\mathbf{T} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -(\mu_1 + \lambda_1) & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -(\mu_2 + \lambda_2) & \lambda_2 & 0 & \cdots \\ 0 & 0 & \mu_3 & -(\mu_3 + \lambda_3) & \lambda_3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (4.10)$$

Con la construcción de un proceso de nacimiento y muerte usando las hipótesis mencionadas al principio queda más claro el porque la matriz de tasas tiene como suma de renglones igual a cero. Otra de las propiedades que estaremos trabajando es que no consideraremos un tamaño de población infinito, sino que este tamaño puede ser una constante  $N$  muy grande, que si la población llega a esta cantidad entonces solo habrá defunciones, por lo que la estructura de la matriz de tasas  $\mathbf{T} \in \mathbb{R}^{(N+1) \times (N+1)}$  de un proceso de nacimiento y muerte finito estará dada por

$$\mathbf{T} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots & 0 \\ \mu_1 & -(\mu_1 + \lambda_1) & \lambda_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & -(\mu_2 + \lambda_2) & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \\ 0 & 0 & \cdots & \mu_{N-1} & -(\mu_{N-1} + \lambda_{N-1}) & \lambda_{N-1} \\ 0 & 0 & \cdots & 0 & \mu_N & -\mu_N \end{bmatrix} \quad (4.11)$$

Los procesos de nacimiento y muerte son importantes al trabajar en la teoría de colas. En la literatura se especifican 6 valores distintos, y se denota como la **notación de Kendall**<sup>2</sup> a

<sup>2</sup>David George Kendall. (1918-2007). Matemático y estadístico inglés. Trabajó en las áreas de probabilidad y teoría de colas.

la sexteta

$$A / B / c / D / E / F$$

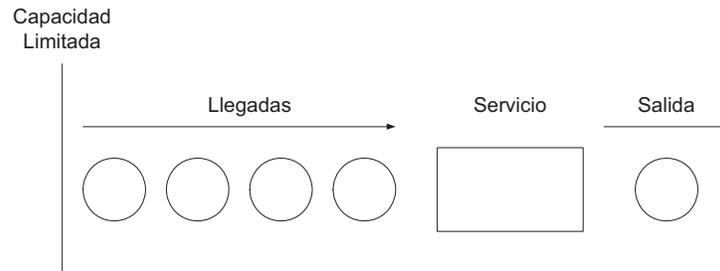
donde el primer valor  $A$  hace referencia al proceso de arribo, el valor  $B$  hace referencia a la distribución del tiempo de servicio, el valor  $c$  hace referencia al número de servidores, el valor  $D$  hace referencia a la capacidad máxima que tiene el sistema, el valor  $E$  describe el tamaño de la población que solicita servicio (que en general es infinito) y por último el valor  $F$  hace referencia a la disciplina de como se atiende en los servidores. Ahora bien, la dupla  $A / B$  puede tomar alguno de los siguiente valores

$M$  significa en inglés "memoryless" por pérdida de la memoria ó del tipo exponencial

$G$  significa una distribución general para el arribo o el tiempo de servicio

$GI$  significa una distribución general independiente en los tiempos de arribo

Un modelo básico pero importante es el dado por el cuarteto  $M / M / 1 / N$ , que tiene como hipótesis de que los arribos son independientes y de un proceso Poisson, los tiempos de servicios tienen una distribución exponencial, solamente hay un sólo servidor y se tiene una capacidad máxima de  $N$  trabajos en espera (junto con el que se está atendiendo). Se asume implícitamente que hay una población infinita y que la modalidad del servicio es FIFO. Una descripción de esta modalidad está dada por el siguiente diagrama



La matriz de tasas  $\mathbf{T}$  asociada a este modelo está dada por

$$\mathbf{T} = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \cdots & 0 \\ \mu & * & \lambda & 0 & \cdots & 0 \\ 0 & \mu & * & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \mu & * & \lambda \\ 0 & 0 & 0 & 0 & \mu & -\mu \end{bmatrix}.$$

donde si se tiene una capacidad máxima de  $N$  clientes en el sistema, entonces se tiene una matriz  $\mathbf{T} \in \mathbb{R}^{(N+1) \times (N+1)}$  y el espacio de estados está dado por  $\mathbb{E} = \{0, 1, 2, \dots, N\}$ . Los valores en la diagonal dados por  $*$  hacen que la suma por renglones sea cero. Se asume en este caso que las tasas son constantes, si fuera un sistema de colas de un super mercado por ejemplo, una cajera no se vería intimidada por la longitud de la cola y estaría atendiendo a los clientes con el mismo tiempo de servicio. También el arribo de los clientes al servidor es constante, que esto nos dice que el tamaño de la cola no influye en la decisión de un cliente para incorporarse a esperar el servicio. En la práctica estas tasas  $\lambda$  y  $\mu$  no necesariamente





```

%
% N = numero clientes en el sistema
% c = numeros servidores (c<=N)
% lam = tasa de arribo promedio
% mu = tasa de servicio promedio
% eta = tiempo promedio espera impaciencia
% Ak = probabilidades de incorporarse a la cola
%      (de la forma [1 1 1 ... 1 Ak(c) Ak(c+1) ... Ak(N-1) ])
%
% FORMA DE USARSE: T=sistemaMMcNgeneral(N,c,lam,mu,eta,Ak)

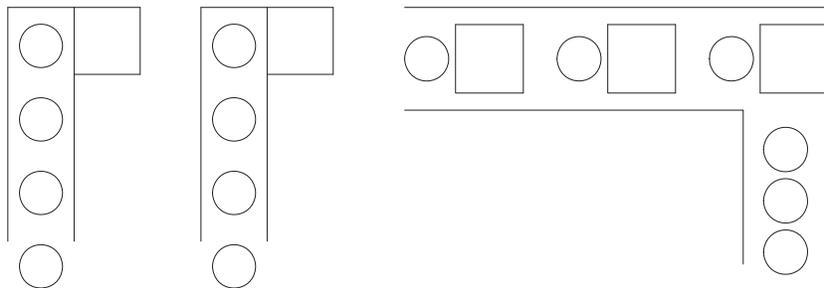
% T e R(N+1,N+1)
% Si no hay probabilidades de que se incorpore a la fila
if( nargin==5 )
    Ak = ones(1,N);
end

w=length(Ak);
if( w~=N )
    error(sprintf(' El vector de probabilidad debe ser de tamaño N = %g ', N))
end

% Se construyen las diagonales superior e inferior
Dsup = lam*Ak; % Diagonal de arribo
vap1 = mu*[1:c , c*ones(1,N-c)]; % Elementos tasa servicio
vap2 = [ 0*[1:c], eta*[1:(N-c)] ]; %Elementos tasa abandono
Dinf = vap1 + vap2;
Dprin = [ Dsup, 0] + [0,Dinf]; % Diagonal principal
T=diag(Dsup,1) + diag(Dinf,-1)+ diag(-Dprin); % Matriz de tasas

```

Con este programa se desea analizar un sistema común en nuestras vida cotidiana, que son las líneas de espera en las tiendas de autoservicio. A continuación se muestra un diagrama de una "mini tienda" de autoservicio.



Esto implica que los dos primeros servidores (de izquierda a derecha) se pueden modelar mediante un modelo  $M / M / 1 / N$ , donde  $N$  es la cola más larga que se puede formar. Cada uno de estos servidores tienen distintas tasas de servicio  $\mu_1$  y  $\mu_2$  (esto porque puede depender de la experiencia de cada cajera) pero las tasas de arribo serían iguales a  $\lambda$ . En el

caso de la configuración de la tercera caja se tendría un modelo  $M / M / c / N$ , con ello se tendría modelado lo que es la sección de "cajas rápidas", por lo que la tasa de servicio  $\mu_3$  tiene que ser más baja comparándola con  $\mu_1$  y  $\mu_2$  pero la tasa de arribo de clientes  $\lambda_2$  tiene que ser más alta. En este caso puede existir una impaciencia del cliente al encontrarse con una fila demasiado larga. Inicialmente nos interesa calcular la probabilidad de que el sistema se encuentra en un estado óptimo (sin clientes). También se puede obtener para cada servidor el tiempo promedio que un cliente tardará en ser atendido. Para poner un caso en particular, asumamos los siguientes datos para los primeros dos servidores

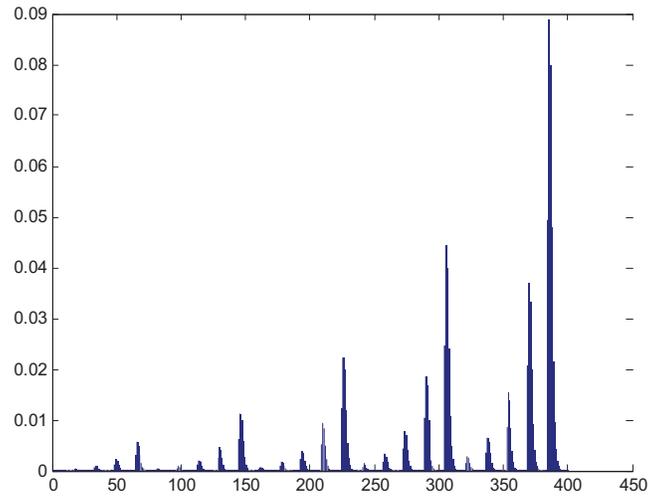
$$\left\{ \begin{array}{l} \lambda = \text{dos clientes cada 5 minutos} \\ \frac{1}{\mu_1} = \text{un cliente cada 5 minutos} \\ \frac{1}{\mu_2} = \text{un cliente cada 6 minutos} \end{array} \right. \implies \left\{ \begin{array}{l} \hat{\lambda} = 24 \text{ clientes por hora} \\ \hat{\mu}_1 = 12 \text{ clientes por hora} \\ \hat{\mu}_2 = 10 \text{ clientes por hora} \end{array} \right.$$

Y los siguientes datos para el tercer servidor

$$\left\{ \begin{array}{l} \lambda_2 = \text{tres clientes cada 5 minutos} \\ \frac{1}{\mu_3} = \text{un cliente cada 3 minutos} \end{array} \right. \implies \left\{ \begin{array}{l} \hat{\lambda}_2 = 36 \text{ clientes por hora} \\ \hat{\mu}_3 = 20 \text{ clientes por hora} \end{array} \right.$$

Con ello se quiere analizar las probabilidades límites. Para ello se construyen las matrices de tasas para cada servidor, por el momento asumiremos que  $N = 4$  para los primeros dos casos y para el tercer caso se tomarán  $c = 4$  y  $N = 15$ .

```
>> %Servidor 1
>> T1=sistemaMMcNgeneral(4,1,24,12,0);
>> %Servidor 2
>> T2=sistemaMMcNgeneral(4,1,24,10,0);
>> %Servidor 3
>> T3=sistemaMMcNgeneral(15,4,36,20,0);
>> %Proceso global
>> T=generador_redes(T1,T2,T3);
>> % Vector estacionario
>> tic, Vest = vector_estacionarioGTH(T); tiempo1=toc
tiempo1 =
    39.1887
```



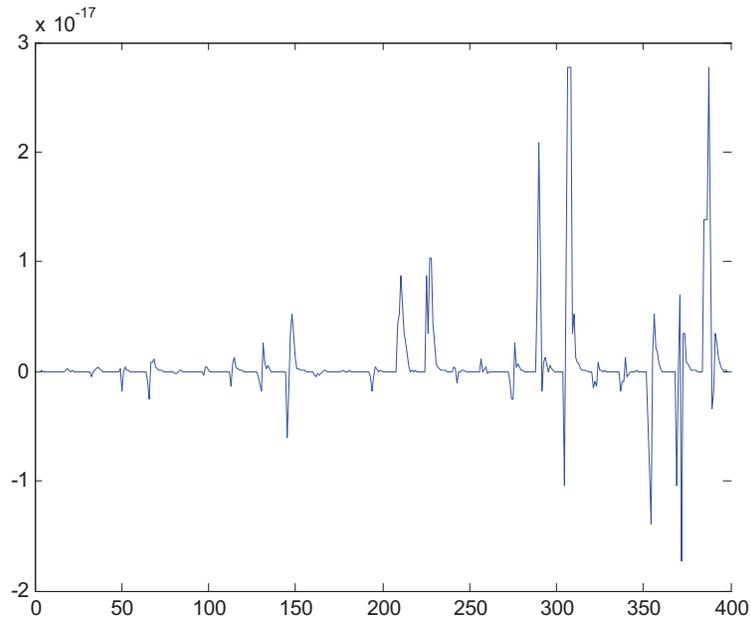
Vector estacionario

De esta forma para encontrar el vector estacionario Mat/Oct se tardó aproximadamente 39.1887 segundos, para obtener que la probabilidad de que los tres servidores desocupado es de  $\pi_{(0,0,0)} = 9.2833 \times 10^{-5}$ . De manera alternativa se puede encontrar este vector estacionario de una forma más práctica sin tener que construir todo el proceso global. Para ello si se tienen  $k$  procesos en paralelo con matrices de tasas  $\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_k$ , entonces el proceso global  $\mathbf{T}$  esta dado por la suma tensorial. Encontrar el vector estacionario de este proceso global es equivalente a encontrar los vectores estacionarios de cada una de las matrices  $\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_k$  dados por  $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \dots, \boldsymbol{\pi}_k$  y después realizar el producto tensorial de estos vectores estacionarios para obtener el vector estacionario del proceso global  $\boldsymbol{\pi}_{\text{Global}} = \boldsymbol{\pi}_1 \otimes \boldsymbol{\pi}_2 \otimes \dots \otimes \boldsymbol{\pi}_k$ . A continuación se muestra la ventaja en tiempo computacional al realizar este procedimiento en Mat/Oct.

```
>> tic, Ve1=vector_estacionarioGTH(T1); Ve2=vector_estacionarioGTH(T2);
    Ve3=vector_estacionarioGTH(T3); Vest2=kron(Ve1,kron(Ve2,Ve3));
    tiempo2=toc;
>> tiempo2
tiempo2 =
    0.1099
```

Es un ahorro en tiempo bastante considerable, y para mostrar que son equivalentes se

muestran los residuales del método 1 contra el método 2.



y con esto también se obtiene que los residuales son muy buenos, ya que son del orden  $10^{-17}$ , y están dentro del límite de la unidad de redondeo de Mat/Oct. Es importante ya que esto se puede aplicar para más servidores cada uno con tiempos de arribo y de servicio distinto. Otra información relevante que se puede obtener de este proceso global es poder estimar los tiempos promedio de cuando alguno de los servidores se desocupen. Otra pregunta importante de la teoría de colas es poder estimar el promedio de clientes en el sistema en estado estable. Para ello se obtiene el vector estacionario de cada servidor y se calcula la cantidad:

$$L = \sum_{k=0}^N k\pi_k .$$

Esta cantidad nos dirá cuántos clientes se espera tenga cada servidor en promedio cuando el sistema se ha estabilizado. Se puede calcular fácilmente esta estadística básica.

```
>> % Para el servidor 1
>> Ls1 = sum([0:4].*Ve1)
Ls1 =
    3.1613

>> % Para el servidor 2
>> Ls2 = sum([0:4].*Ve2 )
Ls2 =
    3.3493

>> % Para el servidor 3
>> Ls3 = sum([0:15].*Ve3)
```

```
Ls3 =
    1.9050
```

Con estos datos se observa que en promedio los servidores 1 y 2 tendrán 3 clientes en promedio, y que el servidor 3 tendrá en promedio 2 clientes en espera. Otro dato que se puede obtener de cada servidor es el tiempo en promedio que cada cliente tiene que esperar para poder recibir el servicio, con ello solo es necesario que hagamos el estado 0 absorbente, los demás estados serán transitorios y con ello se podrá estimar fácilmente el tiempo promedio de espera. Para facilitar el cálculo se hace un programa m-file que permite transformar algún conjunto de estados en absorbentes.

```
function Ta = estados_absorbentes(T,absorbs)

% ESTA FUNCION TRANSFORMA LOS ESTADOS "absorbs"
% EN ABSORBENTES, Y REGRESA UNA NUEVA MATRIZ "Ta"
% PARA NO MODIFICAR LA MATRIZ ORIGINAL.
%
% FORMA DE USARSE: Ta = estados_absorbentes(T,absorbs)

Ta=T;
[n,n]=size(T);

for k=absorbs
    Ta(k,:)=zeros(1,n);
end
```

Y con ayuda de este programa se analizará de manera práctica cuál es el tiempo promedio que un servidor tardará en desocuparse, tomando en cuenta el arribo de clientes al sistema. La función *estadisticasCTMC.m* regresa solamente el vector de tiempos promedio de salida  $\mathbf{M}^*$  y la matriz de probabilidad de última absorción  $\mathbf{F}$ .

```
>> T1=sistemaMMcNgeneral(4,1,24,12,0); T1a=estados_absorbentes(T1,1);
>> [clases1,R1,Q1]=extraccionRQcmtc(T1a);
>> [Mast1,F1]=estadisticasCTMC(Q1,R1); Mast1
Mast1 =
    1.2500
    1.8333
    2.0833
    2.1667

>> T2=sistemaMMcNgeneral(4,1,24,10,0); T2a=estados_absorbentes(T2,1);
>> [clases2,R2,Q2]=extraccionRQcmtc(T2a);
>> [Mast2,F2]=estadisticasCTMC(Q2,R2); Mast2
Mast2 =
    2.2984
```

3.2144  
3.5544  
3.6544

```
>> T3=sistemaMMcNgeneral(15,4,36,20,0); T3a=estados_absorbentes(T3,1);
>> [clases3,R3,Q3]=extraccionRQcmtc(T3a);
>> [Mast3,F3]=estadisticasCTMC(Q3,R3); Mast3'
ans =
```

0.1441	0.1964	0.2267	0.2494	0.2721	0.2948	0.3175	...
0.3402	0.3629	0.3854	0.4077	0.4295	0.4502	0.4683	0.4808

Estos resultados confirman que los servidores 1 y 2 por lo general siempre estarán ocupados, en promedio se tardará en estar sin servicio 2 horas 10 minutos y 3 horas 40 minutos aproximadamente respectivamente. Para el servidor 3 el tiempo de desocupación es mejor, ya que en promedio se tardará 28 minutos en desocuparse, aunque en este caso el tiempo puede aumentar si se permite que la línea de espera crezca, que el tiempo de servicio se haga más lento y por lo tanto la línea de espera se sature. Con esto se aprecia que entre más tiempo de servicio se tarde un servidor la línea de espera será más larga. Con ello se desea mejorar el proceso de poder obtener información conjunta del proceso de cuando al menos un servidor estará desocupado. El problema es que se tienen una gran cantidad de ternas de la forma  $(0, i, j)$  ó  $(i, 0, j)$  ó  $(i, j, 0)$  que representarían cuando el primer, segundo o tercer servidor están desocupados. Lo que se busca es fusionar estas ternas para formar un espacio de estados reducido que facilite la interpretación de resultados. Con ello en Kemeny y Snell [30] definen los elementos para que una cadena de Markov finita  $\{X_n\}_{n \in \mathbb{N}}$  sea truncable<sup>3</sup>, y estas definiciones se extienden a los procesos de Markov finitos  $\{X(t)\}_{t \in \mathbb{R}^+}$ . La idea central de poder truncar un proceso es que se tiene el espacio de estados  $\mathbb{E} = \{1, 2, \dots, n\}$  y se quiere fusionar ciertos estados en un "mega estado"  $\mathbb{E}_j$  para construir un nuevo espacio de estados reducido de la forma  $\mathbb{E}_T = \{\mathbb{E}_1, \dots, \mathbb{E}_m\}$ . Si se cumple que el proceso es truncable, entonces no se perderá información fundamental de la matriz de transición. Para definir este proceso truncable se construyen dos matrices  $\mathbf{U} \in \mathbb{R}^{m \times n}$  y  $\mathbf{V} \in \mathbb{R}^{n \times m}$  de la siguiente manera. Los renglones de la matriz  $\mathbf{U} \in \mathbb{R}^{m \times n}$  cumplen que el  $i$ -ésimo renglón es igual al vector de probabilidad uniforme teniendo entradas iguales para los estados del grupo  $\mathbb{E}_i$  y cero en los demás estados. Las columnas de la matriz  $\mathbf{V} \in \mathbb{R}^{n \times m}$  cumplen que la  $j$ -ésima columna es igual al vector de 1's cuyos componentes corresponden a los estados en  $\mathbb{E}_j$  y cero en otro caso. En Kemeny y Snell [30] se demuestra las condiciones que cumple una cadena de Markov finita para que sea truncable el proceso.

**Teorema 4.9** *Sea  $\{X_k\}_{k \in \mathbb{N}}$  una cadena de Markov con matriz de transición  $\mathbf{P}$  y espacio de estados  $\mathbb{E}$ . Se particiona el proceso con nuevo espacio de estados  $\mathbb{E}_T = \{\mathbb{E}_1, \dots, \mathbb{E}_m\}$  con  $\bigcup_k \mathbb{E}_k = \mathbb{E}$ . Sean  $\mathbf{U}$  y  $\mathbf{V}$  las matrices definidas anteriormente. Entonces el proceso es*

<sup>3</sup>El término inglés es "lumpable", pero no existe una traducción directa al español, por lo que se propone la tradición "truncable", porque se reducirá el espacio de estados original a otro que contenga menos estados.

truncable con respecto a  $\mathbb{E}_T$  si se cumple

$$\boxed{\mathbf{VUPV} = \mathbf{PV}} \quad (4.14)$$

Si el proceso es truncable, entonces la matriz de transición  $\tilde{\mathbf{P}}_T$  del proceso truncable está dado por

$$\boxed{\tilde{\mathbf{P}}_T = \mathbf{UPV}} \quad (4.15)$$

Este resultado se aplica también de manera similar cuando se tiene un proceso de Markov finito. El proceso  $\{X(t)\}_{t \in \mathbb{R}^+}$  es truncable para el espacio de estados  $\mathbb{E}_T$  si  $\mathbf{VUP}(t)\mathbf{V} = \mathbf{P}(t)\mathbf{V}$  para todo  $t \in \mathbb{R}^+$ , y si es truncable entonces la matriz de transición está dada por  $\tilde{\mathbf{P}}_T(t) = \mathbf{UP}(t)\mathbf{V}$ . La misma condición se cumple con la matriz de tasas  $\mathbf{T}$ . Con estos resultados se puede obtener información general del tiempo que se tardará que los servidores se desocupen por primera vez, asumiendo la independencia de cada servidor. En nuestro problema general generamos los siguientes "mega estados"

- $\mathbb{E}_1$  = el servidor 1 está vacío
- $\mathbb{E}_2$  = el servidor 2 está vacío
- $\mathbb{E}_3$  = el servidor 3 está vacío
- $\mathbb{E}_{1,2}$  = los servidores 1 y 2 están vacíos
- $\mathbb{E}_{1,3}$  = los servidores 1 y 3 están vacíos
- $\mathbb{E}_{2,3}$  = los servidores 2 y 3 están vacíos
- $\mathbb{E}_{1,2,3}$  = los tres servidores están vacíos

Con esto se obtiene el nuevo espacio de estados

$$\mathbb{E}_T = \{\mathbb{E}_{1,2,3}, \mathbb{E}_1, \mathbb{E}_2, \mathbb{E}_3, \mathbb{E}_{1,2}, \mathbb{E}_{1,3}, \mathbb{E}_{2,3}, e_1, e_2, \dots, e_m\}$$

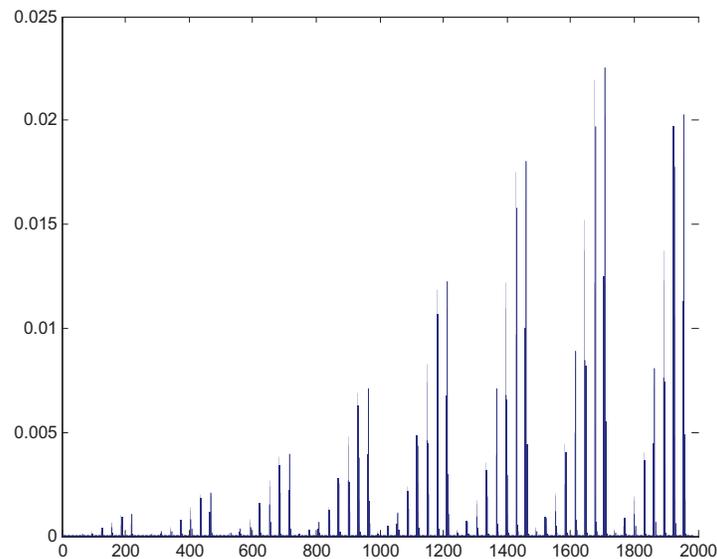
donde los demás estados  $e_j$  representan el estado de los servidores acomodados de forma lexicográfica. Queremos analizar el tiempo promedio que se tarda el sistema global en tener los procesos libres, y para ello los mega estados  $\mathbb{E}_{1,2,3}, \mathbb{E}_1, \mathbb{E}_2, \mathbb{E}_3, \mathbb{E}_{1,2}, \mathbb{E}_{1,3}$  y  $\mathbb{E}_{2,3}$  se convertirán en estados absorbentes. Primero se construye el producto cartesiano  $\{0, 1, \dots, 4\} \times \{0, 1, \dots, 4\} \times \{0, 1, 2, \dots, 15\}$  y se obtendrá un total de 400 ternas, que representan todas las posibles combinaciones posibles de los servidores. Renombrando las ternas en orden lexicográfico  $1^*, 2^*, \dots, 400^*$  se construye un script para analizar este proceso global. Los resultados iniciales muestran que el proceso no es truncable con respecto a esta partición  $\mathbb{E}_T$ , debido a que se obtiene  $\mathbf{VUTV} \neq \mathbf{TV}$ , esto implicaría que estaremos perdiendo cierta información sobre las transiciones de los estados originales, la ventaja es que en teoría sabremos cuál será el mega estado más probable que se visite, que de manera intuitiva deberá ser el mega estado  $\mathbb{E}_3$ , que el tercer servidor se desocupe primero. Se muestran a continuación solamente unos datos obtenidos de este análisis inicial, mostrando algunas entradas de la matriz  $\mathbf{F}$  y

del vector  $\mathbf{M}^*$  (este último multiplicado por un factor de 60 para mostrar los resultados en minutos)

Estado	$\mathbb{E}_{1,2,3}$	$\mathbb{E}_1$	$\mathbb{E}_2$	$\mathbb{E}_3$	$\mathbb{E}_{1,2}$	$\mathbb{E}_{1,3}$	$\mathbb{E}_{2,3}$	$\mathbf{M}_{(60)}^*$
(1, 1, 8)	0	0.3814	0.3073	0.3111	0	0	0	7.9514
(1, 1, 15)	0	0.4058	0.3230	0.2710	0	0	0	10.3416
(2, 2, 8)	0	0.2171	0.1486	0.6341	0	0	0	14.5997
(2, 2, 15)	0	0.2620	0.1761	0.5618	0	0	0	19.5349
(3, 2, 10)	0	0.1361	0.1726	0.6912	0	0	0	17.9743
(3, 3, 10)	0	0.1482	0.0889	0.7627	0	0	0	19.4471
(3, 4, 15)	0	0.1862	0.0845	0.7291	0	0	0	24.1038
(4, 3, 15)	0	0.1492	0.1124	0.7382	0	0	0	24.3427
(4, 4, 15)	0	0.1530	0.0877	0.7591	0	0	0	24.8691

Estos datos inicialmente nos confirman que dos servidores no pueden estar simultáneamente desocupados, por lo que no era necesario fusionar los estados que corresponden a  $\mathbb{E}_{1,2}, \mathbb{E}_{1,3}, \mathbb{E}_{2,3}$  ni tampoco al estado único  $\mathbb{E}_{1,2,3}$ . Otra observación importante es que en general el servidor 3 siempre tiene mejor tiempo de servicio, y se puede mejorar si se tienen más servidores disponibles. Este análisis se puede volver a reconfigurar para analizar las probabilidades de servicio asumiendo que el servidor 1 y 2 mejoran un poco su tiempo de servicio, y que al servidor 3 se le aumenten más servidores. Ahora bien, se analizará a continuación un proceso similar al descrito anteriormente, se tomará ahora en cuenta el factor de impaciencia  $\eta$  que puede tener un cliente para abandonar el sistema si no ha sido atendido, que las líneas de espera aumenten en los tres servidores, pero que en los servidores 1 y 2 exista una probabilidad de que el cliente decida o no incorporarse a la línea de espera, y ver el rendimiento global del sistema. Para este caso de análisis se tomará que un cliente abandonará el sistema si su tiempo promedio es de  $\eta^{-1} = 1.5$  horas, por lo que el factor de impaciencia que se tiene es  $\eta = \frac{2}{3}$ . También se aumentará las líneas de espera de los tres servidores, para los dos primeros casos se tomará de  $N = 7$  pero las probabilidades de incorporarse a la línea de espera serán de  $a_4 = 0.9$ ,  $a_5 = 0.8$  y  $a_6 = 0.6$ . Para el tercer servidor se aumentará a  $c = 5$  procesos en servicio y la línea de espera que aumente hasta  $N = 30$ , en este caso se asume que el cliente no abandona la línea de espera. Con ello poder realizar un análisis similar al que se hizo anteriormente, y observar si existe un mejoramiento en los tiempos de servicio. Este nuevo proceso tendrá una cantidad de  $(7 + 1)(7 + 1)(30 + 1) = 1984$  estados.

```
>> T1=sistemaMMcNgeneral(7,1,24,12,2/3,[1 1 1 1 0.9 0.8 0.6]);
>> T2=sistemaMMcNgeneral(7,1,24,10,2/3,[1 1 1 1 0.9 0.8 0.6]);
>> T3=sistemaMMcNgeneral(30,5,36,20,0);
>> Ve1=vector_estacionarioGTH(T1);
>> Ve2=vector_estacionarioGTH(T2);
>> Ve3=vector_estacionarioGTH(T3);
>> VeTotal=kron(Ve1,kron(Ve2,Ve3));
>> bar(VeTotal)
```



Con esto se observa un comportamiento similar al mostrado en la primer configuración que se analizó previamente. La probabilidad de que el proceso se encuentre desocupado es de  $\pi_{(0,0,0)} = 1.0056 \times 10^{-5}$ , del mismo orden de magnitud que con la configuración original. También en este caso se analizará el tiempo promedio en que un servidor tardará en estar desocupado. Para ello se convierte el estado 0 en absorbente, en los tres casos.

```
>> T1a=estados_absorbentes(T1,1);
>> [clases1,R1,Q1]=extraccionRQcmtc(T1a);
>> [Mast1,F1]=estadisticasCTMC(Q1,R1); Mast1
```

```
Mast1 =
    3.4373
    5.1143
    5.9577
    6.3846
    6.615
    6.7389
    6.8014
```

```
>> T2a=estados_absorbentes(T2,1);
>> [clases2,R2,Q2]=extraccionRQcmtc(T2a);
>> [Mast2,F2]=estadisticasCTMC(Q2,R2); Mast2
```

```
Mast2 =
    8.1263
   11.471
   12.915
   13.556
   13.865
   14.018
   14.089
```

```
>> T3a=estados_absorbentes(T3,1);
>> [clases3,R3,Q3]=extraccionRQcmtc(T3a);
>> [Mast3,F3]=estadisticasCTMC(Q3,R3); [Mast3, 60*Mast3]
ans =
    0.1410    8.4591
    0.1915   11.4919
    0.2199   13.1950
    0.2394   14.3669
    0.2551   15.3044
    0.2707   16.2419
    0.2863   17.1794
    0.3019   18.1169
    0.3176   19.0544
    0.3332   19.9919
    0.3488   20.9294
    0.3644   21.8669
    0.3801   22.8044
    0.3957   23.7419
    0.4113   24.6794
    0.4269   25.6169
    0.4426   26.5544
    0.4582   27.4919
    0.4738   28.4294
    0.4894   29.3669
    0.5051   30.3043
    0.5207   31.2417
    0.5363   32.1790
    0.5519   33.1157
    0.5675   34.0512
    0.5831   34.9830
    0.5984   35.9048
    0.6133   36.7985
    0.6269   37.6145
    0.6369   38.2145
```

Estos datos arrojan como resultados que al crecer la línea de espera de los primeros dos servidores el tiempo que se tardará en desocuparse es muy grande, asumiendo que la tasa de llegada y de servicio se mantiene igual. Como la unidad de tiempo está dado en horas, si los servidores están con su máxima capacidad entonces para que no tenga servicio la línea 1 y 2 es necesario esperar en promedio 6.8 horas y 14.089 horas respectivamente. Para el tercer servidor en general los tiempos de servicio son buenos, aproximadamente se espera que en 38 minutos en promedio se desocupe el tercer servidor. En la práctica las líneas de espera en las tiendas de autoservicio crecen demasiado, los tiempos de servicio varían mucho, por lo que en general se deberá tener cuidado al analizar líneas de espera reales. También puede

aumentar el tiempo debido a que se cierre alguna subestación de este servidor, por lo que los clientes tendrán que trasladarse a otra estación de servicio. Una pregunta inicial es ¿cuál es el tiempo promedio que se tarda el proceso en tener los tres servidores desocupados cuando se tiene las líneas de espera en su máxima capacidad? Se busca analizar la esperanza de la variable aleatoria

$$T_D = \min \{t \in \mathbb{R}^+ : (X_1(t), X_2(t), X_3(t)) = (0, 0, 0)\}$$

y como se realizó anteriormente el estado 1\* de nuestro espacio de estados se hace absorbente. Con ello obtenemos como respuesta no tan gratificante lo siguiente

```
>> Tglobal=generador_redes(T1,T2,T3);
>> TglobalA=estados_absorbentes(Tglobal,1);
>> [clasesg,Rg,Qg]=extraccionRQcmtc(TglobalA);
>> [MastG,FG]=estadisticasCTMC(Qg,Rg);
>> MastG(end)
ans =
1517.9
```

Por lo que en la práctica no se verá que los tres servidores se desocupen de manera conjunta hasta que "transcurran" 1518 horas aproximadamente, que en la realidad pues esto es imposible debido a las limitaciones físicas implícitas. Se realizará un análisis similar al construir un nuevo espacio de estados y truncar el proceso con la siguiente estructura

$$\mathbb{E}_T = \{\mathbb{E}_{1,2,3}, \mathbb{E}_1, \mathbb{E}_2, \mathbb{E}_3, e_1, e_2, \dots, e_m\}$$

donde  $\mathbb{E}_{1,2,3}$  representa el estado de que los tres servidores están desocupados, el mega estado  $\mathbb{E}_1$  representa los estados donde el servidor 1 está desocupado, y de manera similar los estados  $\mathbb{E}_2$  y  $\mathbb{E}_3$ . Los demás estados  $e_1, e_2, \dots, e_m$  representan las ternas faltantes. Al realizar esta partición del espacio de estados original se observa de nueva cuenta que el proceso no es truncable con respecto a este espacio de estados, por lo que se está perdiendo información básica sobre la transición entre estados. Algunos resultados básicos de la matriz de última absorción y el vector de tiempos promedios de salida se muestran a continuación.

Estados	$\mathbb{E}_{1,2,3}$	$\mathbb{E}_1$	$\mathbb{E}_2$	$\mathbb{E}_3$	$\mathbf{M}_{(60)}^*$
(0, 0, 1)	0.2210	0.3894	0.3894	0	0.9737
(0, 1, 0)	0.1096	0.5342	0	0.3561	0.8903
(1, 1, 1)	0	0.2582	0.2119	0.5298	3.7783
(1, 1, 30)	0	0.4062	0.3245	0.2692	13.0849
(1, 2, 10)	0	0.4524	0.1081	0.4393	10.4233
(2, 2, 8)	0	0.1973	0.1394	0.6632	13.2915
(2, 2, 30)	0	0.2573	0.1755	0.5670	25.3593
(3, 3, 20)	0	0.1388	0.0842	0.7768	24.9931
(3, 7, 30)	0	0.1688	0.0214	0.8097	33.6336
(4, 3, 30)	0	0.1057	0.0999	0.7972	33.3795
(4, 4, 20)	0	0.0851	0.0461	0.8686	27.0918
(5, 6, 10)	0	0.0344	0.0105	0.9549	19.4269
(6, 4, 1)	0	0.0075	0.0124	0.9800	8.2265
(7, 7, 30)	0	0.0527	0.0242	0.9230	36.9474

Y estos datos confirman que en general los primeros dos servidores es muy difícil que se desocupen rápidamente, y es el tercer servidor el que tiene la mejor tasa de servicio. Por esto es recomendable que las tiendas de autoservicio aumenten la cantidad de cajas para atender a sus clientes, ya que con ello la tasa de arribo a cada servidor disminuiría drásticamente, y se pueden reducir considerablemente los tiempos de espera de los clientes. Aquí se considera una tasa de impaciencia pequeña (promedio largo) pero en la realidad esta tasa se puede ver afectada cuando un cliente ve líneas de espera demasiado grandes y en general abandonar el sistema. Aunque en este caso analizar la teoría de colas con elementos matriciales resulta computacionalmente más intenso, la ventaja que se obtiene es que se puede obtener más información sobre el proceso, se pueden considerar mucho más servidores con tasas de servicio, entre otros, permite obtener respuestas más prácticas que si analiza solamente un solo servidor.

## 4.3 El problema de inclusión

### 4.3.1 El logaritmo de una matriz

En esta parte se hace un análisis particular sobre la movilidad social y la construcción de una matriz de tasas para un proceso a tiempo continuo a partir de una matriz de transición empírica a tiempo discreto, aunque el origen de este estudio teórico se dió al considerar las transiciones de créditos de empresas, y se puede revisar más a fondo en Jarrow et. al [57]. La idea central es construir una matriz de tasas a tiempo continuo a partir de una matriz de transición a tiempo discreto, que en inglés se le conoce como "the embedding problem", y que se puede traducir al español como "el problema de inclusión" o "el problema de incrustación". El problema matemático que se planteó originalmente es el siguiente:

Dada una cadena de Markov  $\{X_\tau\}_{\tau \in \mathbb{N}}$  ¿será posible construir un proceso de Markov  $\{\tilde{X}(\tau)\}_{\tau \in \mathbb{R}}$  tal que la distribución de probabilidad de  $\tilde{X}(\tau)$  al tiempo  $\tau \in \mathbb{N}$  sea igual a la distribución de  $X_\tau$ ?

Se ha estudiado anteriormente que si el proceso de Markov tiene una matriz de tasas  $\mathbf{T}$  entonces las probabilidades de transición están dadas por

$$\mathbf{P}(t) = e^{\mathbf{T}t} \text{ con } t \geq 0$$

entonces, si se toma el tiempo  $t = 1$  obtenemos

$$\mathbf{P} = \mathbf{P}(1) = e^{\mathbf{T}}$$

Esta matriz  $\mathbf{P}$ , se puede pensar como que es la matriz de transición en una etapa de una cadena de Markov discreta, entonces, si quisieramos construir cuál es su matriz de tasas correspondientes, solamente "tendríamos" que despejar la función exponencial

$$\mathbf{P} = e^{\mathbf{T}} \implies \mathbf{T} = \ln(\mathbf{P}) \quad (4.16)$$

El problema con el que se han encontrado en Israel et. al [58] es que en general no se puede construir una matriz de tasas válida para un proceso a tiempo continuo a partir de solamente encontrar el logaritmo de una matriz de transición (si es que existe). Al querer encontrar el logaritmo de una matriz de transición pueden suceder al menos tres casos posibles:

- a) La matriz logaritmo no existe
- b) La matriz logaritmo existe pero no cumple las condiciones de una matriz de tasas ( $\lambda_{jj} \leq 0$  y  $\lambda_{i,j} \geq 0$  para  $i \neq j$ )
- c) La matriz logaritmo existe y cumple las condiciones de una matriz de tasas

Por la expansión en series de Taylor de la función logaritmo se tiene que

$$\ln(x) = (x - 1) - \frac{(x - 1)^2}{2} + \frac{(x - 1)^3}{3} - \frac{(x - 1)^4}{4} + \dots \text{ para } |x - 1| < 1$$

y recordando la expresión (2.5) se obtiene una condición necesaria para la existencia de la matriz logaritmo

$$\ln(\mathbf{P}) = (\mathbf{P} - \mathbf{I}) - \frac{(\mathbf{P} - \mathbf{I})^2}{2} + \frac{(\mathbf{P} - \mathbf{I})^3}{3} - \frac{(\mathbf{P} - \mathbf{I})^4}{4} + \dots \text{ para } \rho(\mathbf{P} - \mathbf{I}) < 1 \quad (4.17)$$

Entonces, con esto, lo que se debe garantizar para la existencia de la matriz logaritmo usando serie de potencias es que  $\rho(\mathbf{P} - \mathbf{I}) < 1$ , es el primer elemento que se debe verificar para construir el logaritmo con este procedimiento. Como primer caso, se tiene la siguiente matriz de transición, que se tomó de Singer y Spilerman [60]

$$\mathbf{P}_1 = \begin{bmatrix} 0.6 & 0.33 & 0.07 \\ 0.302 & 0.56 & 0.138 \\ 0.38 & 0.04 & 0.58 \end{bmatrix}$$

cuyo radio espectral está dado por  $\rho(\mathbf{P}_1) = 0.630095$ , por lo que la serie (4.17) converge. Haciendo cálculos con 100 términos de la serie se obtiene que

$$\mathbf{T}_1 = \ln(\mathbf{P}_1) = \begin{bmatrix} -0.6923 & 0.6393 & 0.0529 \\ 0.4961 & -0.7327 & 0.2367 \\ 0.7068 & -0.1442 & -0.5626 \end{bmatrix}$$

con lo que se observa inicialmente que si existe la matriz logaritmo pero que de las condiciones de una matriz de tasas solamente se cumple aproximadamente que la suma por renglones es cero, pero no se cumple que las entradas que no están en la diagonal sean todas positivas. Un código básico de Mat/Oct se muestra a continuación

```
function T=logaritmo_matriz_serie(P,m)
% ESTA FUNCION CALCULA EL LOGARITMO NATURAL DE UNA MATRIZ
% DE TRANSICION "P" USANDO LA SERIE DE POTENCIAS CON "m"
% TERMINOS
```

```

%
% FORMA DE USARSE: T=logaritmo_matriz_serie(P,m)

n=size(P);
if( n(1)~=n(2) )
    error(' LA MATRIZ NO ES CUADRADA!!! ')
end
n=max(n);

% Se construye la matriz de apoyo
Apoyo=P-eye(n);
% Se calcula el radio espectral de la matriz
RE=max(abs(eig(Apoyo)));

if( RE>=1 )
    error(sprintf(' EL RADIO ESPECTRAL DE LA MATRIZ ES MAYOR...
... QUE 1, QUE ES %g ', RE))
end

T=zeros(n); Apoyo = eye(n); Q=P-eye(n);

for k=1:m
    Apoyo=Apoyo*Q;
    T = T + ( power(-1,k+1)/k )*Apoyo;
end

disp(sprintf(' RADIO ESPECTRAL MENOR QUE 1, DADO POR %g ', RE))

```

Como ejemplo del código se toma la siguiente matriz de transición

```

>> P=[0.9 0.08 0.02; 0.1 0.80 0.1; 0 0 1];
>> T=logaritmo_matriz_serie(P,100)
RADIO ESPECTRAL MENOR QUE 1, DADO POR 0.25247
T =
    -0.1107    0.0946    0.0162
     0.1182   -0.2289    0.1107
         0         0         0

```

Un problema grave que tiene el utilizar esta serie de potencia es los signos alternantes, que siempre se han caracterizado por tener cancelaciones graves, por lo que podría no resultar una opción viable a la hora de estimar la matriz logaritmo. Para ello se puede trabajar con una serie que no tiene signos alternantes y eso garantiza una mejor estabilidad numérica

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \dots = -\sum_{k=1}^{\infty} \frac{x^k}{k} \text{ con } |1-x| < 1 \text{ y } x \neq 1$$

y usando los resultados (2.5) y el hecho de que  $a = 1 - (1 - a)$  se obtiene la siguiente serie matricial para la construcción de la matriz logaritmo

$$\ln(\mathbf{P}) = - \sum_{k=1}^{\infty} \frac{(\mathbf{I} - \mathbf{P})^k}{k} \text{ con } \rho(\mathbf{I} - \mathbf{P}) < 1 \text{ y } \mathbf{P} \neq \mathbf{I} \quad (4.18)$$

Estas dos series (4.17) y (4.18) son formas iniciales para encontrar el logaritmo de una matriz de transición, pero exigen la condición (de momento restrictiva) que  $\rho(\mathbf{P} - \mathbf{I})$  o  $\rho(\mathbf{I} - \mathbf{P})$  sean menores que 1, y esta condición puede ser que no la cumplan algunas matrices de transición pero que si tengan matriz logaritmo. Como segundo ejemplo de una matriz de transición y que existe el logaritmo, dada en Singer y Spilerman [60] pero que su radio espectral es mayor que 1, es la siguiente

$$\mathbf{P}_2 = \begin{bmatrix} 0.260 & 0.169 & 0.248 & 0.323 \\ 0.327 & 0.275 & 0.146 & 0.252 \\ 0.269 & 0.346 & 0.232 & 0.153 \\ 0.162 & 0.285 & 0.305 & 0.248 \end{bmatrix} \quad \rho(\mathbf{P}_2) = 1.02341 \quad (4.19)$$

sin embargo la matriz logaritmo de  $\mathbf{P}_2$  está dada por

$$\mathbf{T}_2 = \begin{bmatrix} -1.6980 & 0.0308 & 0.0251 & 1.6420 \\ 1.5687 & -1.6502 & 0.0548 & 0.0267 \\ 0.0566 & 1.7763 & -1.8480 & 0.0151 \\ 0.0114 & 0.0883 & 1.6488 & -1.7486 \end{bmatrix}$$

Con esto se aprecia que deben existir otras condiciones que se deben de cumplir para que una matriz de transición se le pueda calcular el logaritmo.

### 4.3.2 Condiciones de inclusión

Distintos autores han estudiado y encontrado elementos teóricos que debe de cumplir una matriz de transición para que se pueda construir una matriz de tasas válida, y que se muestran a continuación, y estas condiciones han sido analizadas por distintos autores.

**Teorema 4.10** *Sea  $\mathbf{P} \in \mathbb{R}^{n \times n}$  una matriz de transición y sea  $\mathbf{T}$  una matriz de tasas válida para  $\mathbf{P}$ , es decir,  $e^{\mathbf{T}} = \mathbf{P}$ , entonces se deben cumplir las siguientes condiciones*

1. (Kingman)  $\det(\mathbf{P}) > 0$
2. (Israel)  $\det(\mathbf{P}) \leq \prod P_{j,j}$
3. (Elfving) Ningún otro valor propio distinto de  $\lambda = 1$  debe cumplir  $|\lambda| = 1$
4. (Chung) Para todo par de estados  $i, j$  se debe cumplir que si  $i \rightarrow j$  entonces  $P_{i,j} > 0$

5. (Runnenberg) Todos los valores propios de  $\mathbf{P}$  deben estar contenidos dentro de la región compleja  $z = x(v) + iy(v)$  con  $v \in \left(0, \frac{\pi}{\sin(2\pi/n)}\right)$

$$\begin{aligned}x(v) &= \exp\left(-v + v \cos\left(\frac{2\pi}{n}\right)\right) \cos\left(v \sin\left(\frac{2\pi}{n}\right)\right) \\y(v) &= \exp\left(-v + v \cos\left(\frac{2\pi}{n}\right)\right) \sin\left(v \sin\left(\frac{2\pi}{n}\right)\right)\end{aligned}$$

y su simetría sobre el eje real

6. (Singer y Spilerman) Si la matriz  $\mathbf{P}$  tiene todos sus valores propios distintos y si  $\mathbf{T}$  es una matriz de tasas para  $\mathbf{P}$ , entonces los valores propios de  $\mathbf{T}$  satisfacen

$$|\operatorname{Im} \lambda| \leq |\ln(\det(\mathbf{P}))|$$

7. (Singer y Spilerman) Si la matriz  $\mathbf{P}$  tiene valores propios negativos, entonces no existe una matriz real  $\mathbf{T}$  tal que  $e^{\mathbf{T}} = \mathbf{P}$

Estas son algunas de las propiedades que los autores antes citados han obtenido para el problema de la inclusión, lo que muestra es que se tienen muchas restricciones para que una matriz a tiempo discreto pueda ser incluida en un proceso a tiempo continuo. Para el caso 7) se especifica en Singer y Spilerman [60] que si se tienen valores propios negativos repetidos con multiplicidad par, es posible encontrar una solución al problema de inclusión. Se hará un pequeño m-file en Mat/Oct donde se puedan visualizar algunas de estas restricciones, y verificar en que casos se puede obtener el logaritmo de la matriz  $\mathbf{P}$

```
function region_critica_inclusion(P)

% ESTA FUNCION GRAFICA LA REGION CRITICA DE LA MATRIZ
% DE TRANSICION, PARA VERIFICAR LA INCLUSION A UN
% PROCESO DE MARKOV CONTINUO, USANDO EL CRITERIO DE
% RUNNENBERG
%
% FORMA DE USARSE:  region_critica_inclusion(P)

n=size(P);
if( n(1)~= n(2) )
    error(' LA MATRIZ DEBE SER CUADRADA!!! ')
end

n=max(n);

theta=linspace(0,2*pi,1000); xcirc=cos(theta); ycirc=sin(theta);
v=linspace(0,pi/sin(2*pi/n),1000);
```

```

xv=exp(-v + v*cos(2*pi/n)).*cos(v*sin(2*pi/n));
yv=exp(-v + v*cos(2*pi/n)).*sin(v*sin(2*pi/n));

A=eig(P); Xval=real(A); Yval=imag(A);

hold on
plot(xv,yv,'r','LineWidth',2.2), plot(xv,-yv,'r','LineWidth',2.2)
plot(xcirc,ycirc,'k','LineWidth',2.2)

detP=det(P); PDiag=prod(diag(P));

for k=1:n
    plot(Xval(k),Yval(k),'b*')
end

title(' REGION DE INCLUSION DE RUNNENBERG ')
xlabel(sprintf(' det(P) = %g , Prod(diag) = %g ',detP, PDiag))
if( detP<=0 )
    warning(' DETERMINANTE DE LA MATRIZ ES NEGATIVO')
end

if( detP>=PDiag )
    warning(' EL DETERMINANTE ES MAYOR QUE EL PRODUCTO DE LA DIAGONAL!!! ')
end

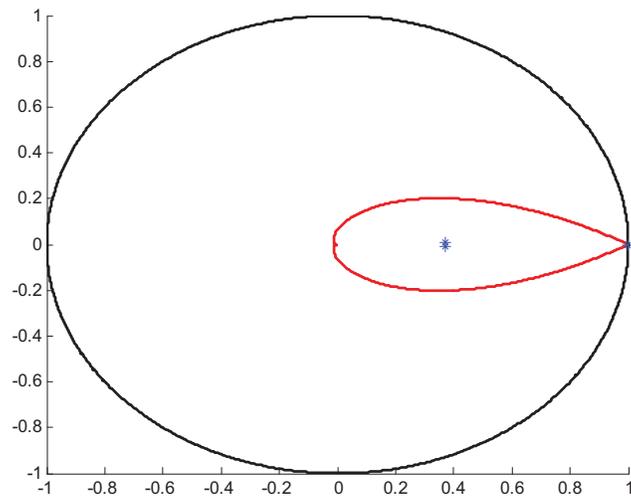
axis([-1 1 -1 1])

hold off

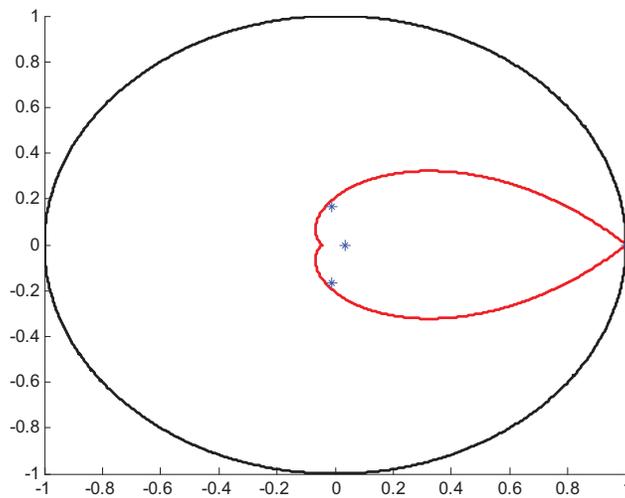
    Con esto analicemos las dos primeras matrices  $\mathbf{P}_1$  y  $\mathbf{P}_2$ 

>> P1=[0.6 0.33 0.07; 0.302 0.56 0.138; 0.38 0.04 0.58];
>> region_critica_inclusion(P1)

```



```
>> % Esta es la matriz donde RE(P-I)>1
>> P2=[0.260 0.169 0.248 0.323; 0.327 0.275 0.146 0.252;
        0.269 0.346 0.232 0.153; 0.162 0.285 0.305 0.248];
>> region_critica_inclusion(P2)
```



El criterio de Runnenberg [67] da una excelente guía visual para poder determinar si existe la matriz logaritmo de una matriz de transición, ya que con esto se puede complementar la idea de que no necesariamente una matriz de transición debe cumplir que  $\rho(\mathbf{P} - \mathbf{I}) < 1$  para que existe la matriz logaritmo, el ejemplo anterior da cuenta de este hecho. Ya que se determinó si los valores propios caen dentro de la región de inclusión, lo que procede es encontrar la matriz logaritmo sin tener que aplicar la serie (4.17). En Israel et. al [58] se

propone utilizar la interpolación de Lagrange dada por

$$g(\mathbf{A}) = \sum_j f(z_j) \prod_{k \neq j} \frac{\mathbf{A} - z_k \mathbf{I}}{z_j - z_k} \quad (4.20)$$

donde  $z_k$  son los valores propios distintos de  $\mathbf{P}$ , y la función  $f(z_j)$  está dada por<sup>4</sup>

$$f(z_j) = \text{Log } |z_j| + i(\text{Arg}(z_j) + 2\pi k_j)$$

con  $k_j = 0, \pm 1, \pm 2, \dots$  y que además se debe cumplir  $|\text{Arg}(z_j) + 2\pi k_j| \leq |\ln(\det(\mathbf{P}))|$ . En Meyer [1] se demuestra que toda función de una matriz  $\mathbf{A}$  puede ser expresado como un polinomio en  $\mathbf{A}$ , y que los elementos  $\mathbf{A}_j = \prod_{k \neq j} \frac{\mathbf{A} - z_k \mathbf{I}}{z_j - z_k}$  se les conoce como los **proyectores**

**espectrales**. Para el caso de que no se tengan valores propios distintos existe un procedimiento para encontrar la función de una matriz, pero para ello es necesario poder encontrar de manera numérica los vectores propios generalizados, y que en Golub y Wilkinson [43] demuestran que es un problema numéricamente inestable, por lo que solamente se trabajará con el caso diagonalizable. Una implementación del algoritmo, al cuál denotaremos por *matriz\_generadora\_IRW.m*, se muestra a continuación

```
function TQ=matriz_generadora_IRW(P)

% ESTA FUNCION IMPLEMENTA EL ALGORITMO PROPUESTO POR
% ISRAEL, ROSENTHAL Y WEI (IRW) PARA LA BUSQUEDA DE
% UNA MATRIZ GENERADORA "TQ" TAL QUE CUMPLA TQ=ln(P)
% HACIENDO USO DE LA INTERPOLACION DE LAGRANGE
% (LA MATRIZ "TQ" PUEDE SER MULTIDIMENSIONAL)
%
% FORMA DE USARSE: T=matriz_generadora_IRW(P)

% Verificamos que la matriz sea cuadrada
n=size(P);
if( n(1)~=n(2) )
    error(' LA MATRIZ DEBE SER CUADRADA!!! ')
end
n=max(n);

% Calculamos los valores propios de la matriz
ValProp=eig(P);

% Verificamos que sean distintos
d=vals_distintos(ValProp);
```

---

<sup>4</sup>Arg( $z$ ) = arctan  $\left(\frac{b}{a}\right)$  es el **argumento principal** de un número complejo  $z = a + bi$

```

if( d==0 )
    error(' NO TODOS LOS VALORES PROPIOS SON DISTINTOS!!! ')
end

detP=det(P);
if( detP<=0 )
    error(' NO EXISTE UNA MATRIZ GENERADORA!!! ')
end

% El valor critico del determinante
VCrit=abs(log(detP));

% Construccion valores propios "generalizados" y calculamos
% la cantidad total de posibles generadoras
TT=1;
for k=1:n
    FRx_k = indices_valor_propio(ValProp(k),VCrit);
    VPFam{k}= FRx_k; TT=TT*length(FRx_k);
end

% Todas las posibles combinaciones
casos=cartesianProduct(VPFam); TQ=zeros(n,n,TT); ydatos=zeros(n,1);

for k=1:TT
    ydatos=casos(k,:);
    TQ(:,:,k)=interpolacion_lagrange_matriz(P,ValProp,ydatos);
end

% La ultima matriz se tomara con la rama principal
TQ(:,:,TT+1)=interpolacion_lagrange_matriz(P,ValProp,log(ValProp));
%-----
%-----

function d=vals_distintos(xdatos)

% VERIFICA DE MANERA RUDIMENTARIA SI TODOS
% LOS VALORES "xdatos" SON DISTINTOS
% d=0 - ALGUN VALOR ES IGUAL
% d=1 - TODOS LOS VALORES SON DISTINTOS

n=length(xdatos); d=1;
for i=1:n
    for j=(i+1):n
        if( xdatos(i)==xdatos(j) )
            d=0;
        end
    end
end

```

```

        end
    end
end

%-----
%-----

function FRx=indices_valor_propio(Vx,VC)
% VC - valor critico del determinante
% Vx - valor propio Vx

ReX = real(Vx);    ImX = imag(Vx);

indices=[0 1 -1 2 -2 3 -3] % Se tomaran 7 posibles valores de las ramas

% Calculamos valores auxiliares
ArgX=atan2(ImX,ReX);    AbsX=sqrt(power(ReX,2)+power(ImX,2));

% Buscamos que ramas cumplen las propiedades
Prueba= abs( ArgX + 2*pi*indices);    ind=find(abs(Prueba)<=abs(VC));

%Todos los posibles valores logaritmicos
FRx = log(AbsX) + sqrt(-1)*Prueba(ind);

%-----
%-----

function result = cartesianProduct(sets)

% Funcion para el producto cartesiano tomada de
% http://stackoverflow.com/questions/4165859/matlab-generate-all-possible-
% combinations-of-the-elements-of-some-vectors
c = cell(1, numel(sets));
[c{:}] = ndgrid( sets{:} );
result = cell2mat( cellfun(@(v)v(:), c, 'UniformOutput',false) );

%-----
%-----

```

Dentro de esta implementación es importante tomar en cuenta que en general Mat/Oct no encontrará de manera exacta los valores propios de una matriz usando el comando "eig",

sino una aproximación. La siguiente matriz

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 1/2 & -1/2 \\ 1/2 & 3/2 & 0 & 0 \\ 1/2 & -1/2 & 2 & 0 \\ 0 & 0 & 1/2 & 3/2 \end{bmatrix}$$

tiene polinomio característico  $p(\lambda) = (\lambda - 1)(\lambda - 2)^3$ , por lo que tiene solamente dos valores propios  $\lambda_1 = 1$  y  $\lambda_2 = 2$ . Haciendo los cálculos en doble precisión se obtiene

```
>> A=[2 0 1/2 -1/2; 1/2 3/2 0 0;
      1/2 -1/2 2 0; 0 0 1/2 3/2];
>> format long g
>> d=eig(A)
d =
           1
    1.99999997640326
    2.00000002359674
           2
```

Por lo que en general hay que tener cierta cautela al trabajar con el comando "eig" de Mat/Oct. Regresando al problema de inclusión, recordemos que la matriz (4.19) tiene un radio espectral mayor que 1, por lo que no se puede utilizar ninguna de las dos series (4.17) ni (4.18), entonces ya se verificó que cumple la condición de Runnenberg, entonces utilicemos el código anterior para encontrar el logaritmo de esta matriz.

```
>> P2=[0.260 0.169 0.248 0.323; 0.327 0.275 0.146 0.252;
       0.269 0.346 0.232 0.153; 0.162 0.285 0.305 0.248];

>> T2=matriz_generadora_IRW(P2);
>> size(T2)
ans =
     4     4    37

>> % La ultima matriz es la rama principal
>> MT=T2(:, :, 37);
>> real(MT) % Solamente se toma la parte real
ans =
   -1.6980    0.0308    0.0251    1.6420
    1.5687   -1.6502    0.0548    0.0267
    0.0566    1.7763   -1.8480    0.0151
    0.0114    0.0883    1.6488   -1.7486
```

Con esto se observa que efectivamente se obtiene el logaritmo de la matriz que se reporta en [60]. Una de las "desventajas" iniciales del código es que genera todas las matrices  $g(\mathbf{A})$  tal que los valores propios de  $\mathbf{A}$  satisfacen la relación  $|\text{Arg}(z_j) + 2\pi k_j| \leq |\ln(\det(\mathbf{P}))|$ , es decir,

construye matrices complejas  $\mathbf{Z}$  tales que  $e^{\mathbf{Z}} = \mathbf{P}$ , pero que no tienen ningún significado práctico en el contexto de los proceso de Markov. Dependiendo del tamaño de la matriz se puede modificar el número de ramas de la función logaritmo compleja y reducir el costo computacional correspondiente.

### 4.3.3 Ajustes a la matriz logaritmo

Con los elementos anteriores se ha visto que en caso de existir la matriz logaritmo como poder encontrar esta matriz con al menos tres procedimientos básicos, pero existe el inconveniente de que no se obtenga una matriz de tasas válidas. Se analizó un ejemplo anteriormente, pero se muestra otra matriz de transición que si existe la matriz logaritmo pero que no cumple las condiciones de una matriz de tasas válida.

$$\mathbf{P} = \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0 & 0.7 & 0.3 \\ 0.1 & 0.1 & 0.8 \end{bmatrix} \quad \ln(\mathbf{P}) = \begin{bmatrix} -0.5123 & 0.4611 & 0.0512 \\ -0.0318 & -0.3763 & 0.4081 \\ 0.1466 & 0.1042 & -0.2509 \end{bmatrix}$$

En caso de que sea absolutamente necesario construir a partir de esta matriz  $\mathbf{T} = \log(\mathbf{P})$  una matriz de tasas válida  $\tilde{\mathbf{T}}$ , existen al menos tres métodos distintos. El primero de ellos fue propuesto por Jarrow et. al [57], donde analizan las probabilidades de transición de compañías a distintas tasas de riesgo de crédito. Para poder estimar una matriz de tasas asumen ellos que las probabilidades de que una compañía cambie de situación financiera más de una vez durante un año son pequeñas. También en su modelo de construcción de tasas asumen que existe un estado absorbente que representa la clasificación más baja a la que una compañía puede acceder, que es de insolvencia económica. La matriz de tasas asociada no depende del cálculo de la matriz logaritmo, por lo que es un procedimiento relativamente sencillo de obtener. La matriz de tasas  $\mathbf{T}_{JLT}$  propuesta en Jarrow et. al [57] está dada por

$$\begin{aligned} \lambda_{i,i} &= \ln(P_{i,i}) \\ \lambda_{i,j} &= \frac{P_{i,j} \ln(P_{i,i})}{P_{i,i} - 1} \quad \text{con } i \neq j \end{aligned} \quad (4.21)$$

Este procedimiento asume que  $P_{i,i} > 0$ . El código básico en Mat/Oct para la construcción de la matriz de tasas la llamaremos *generador\_JLT.m*

```
function T=generador_JLT(P)

% ESTA FUNCION CONSTRUYE UNA MATRIZ DE TASAS VALIDA
% PARA LA MATRIZ "P" USANDO EL METODO DE JARROW-LANDO-TURNBULL
%
% FORMA DE USARSE: Q=generador_JLT(P)

n=size(P);
```

```

if( n(1)~=n(2) )
    error(' LA MATRIZ DE TRANSICION DEBE SER CUADRADA!!! ')
end
n=max(n); T=zeros(n);

% Ningun elemento de la diagonal debe ser igual a cero
if(prod(diag(P))==0)
    error('NO EXISTE MATRIZ GENERADORA POR EL METODO JLT!!! ')
end

for k=1:n
    if(P(k,k)==1)
        T(k,:)=zeros(1,n);
    else
        valor=log(P(k,k));
        T(k,:)=P(k,:)*valor/(P(k,k)-1);
        T(k,k)=valor;
    end
end
end

```

Regresando a la matriz inicial se puede obtener fácilmente la matriz de tasas  $JLT$  correspondiente

```

>> P1=[0.6 0.3 0.1; 0 0.7 0.3; 0.1 0.1 0.8];
>> T=generador_JLT(P1)
T =
    -0.5108    0.3831    0.1277
         0    -0.3567    0.3567
    0.1116    0.1116   -0.2231

>> expT=expm(T)
expT =
    0.6064    0.2561    0.1375
    0.0139    0.7164    0.2697
    0.0783    0.0993    0.8224

>> norm(P1 - expT,2)
ans =
    0.0689

```

Este ejemplo básico muestra que al calcular la matriz exponencial de la matriz  $JLT$  no se obtiene exactamente la matriz original, sino una matriz aproximada, esto en algunos casos podría no ser aceptable no poder tener la matriz original. Otras modificaciones se obtienen de primero obtener la matriz logaritmo y después se modifican los renglones de esta matriz

de manera tal que se obtenga una matriz de tasas válida. Estas dos formas de ajustar a la matriz están descritas en Israel et. al [58], y se muestran a continuación, donde  $\tilde{\mathbf{T}}$  es la matriz ajustada y  $\mathbf{T}$  es la matriz original.

### Método 1

$$\begin{aligned}\tilde{\lambda}_{i,j} &= \max(\lambda_{i,j}, 0) \text{ para } i \neq j \\ \tilde{\lambda}_{i,i} &= \lambda_{i,i} + \sum_{j \neq i} \min(\lambda_{i,j}, 0)\end{aligned}$$

Lo que nos dice este primer método de ajuste es que se buscan los elementos fuera de la diagonal que sean menores que cero, se sustituyen por el valor de cero y después se actualiza los valores de la diagonal para que la suma por renglones sea cero. El segundo método propuesto es el siguiente

### Método 2

$$\begin{aligned}G_i &= |\lambda_{i,i}| + \sum_{j \neq i} \max(\lambda_{i,j}, 0) \\ B_i &= \sum_{j \neq i} \max(-\lambda_{i,j}, 0) \\ \tilde{\lambda}_{i,j} &= \begin{cases} 0 & \text{si } i \neq j \text{ y } \lambda_{i,j} < 0 \\ \lambda_{i,j} - \frac{B_i |\lambda_{i,j}|}{G_i} & \text{si } G_i > 0 \\ \lambda_{i,j} & \text{si } G_i = 0 \end{cases}\end{aligned}$$

Lo que hace el segundo método es sumar todos los valores negativos del mismo renglón en todos los elementos de este, y que esta suma sea proporcional en valor absoluto. Una tercera opción para ajustar a la matriz logaritmo es analizada en Inamura [71], que de manera similar primero transforma los elementos negativos fuera de la diagonal a cero y después hace un ajuste ponderado con los elementos faltantes.

### Método 3

$$\begin{aligned}\tilde{\lambda}_{i,j} &= \begin{cases} 0 & \text{si } i \neq j \text{ y } \lambda_{i,j} < 0 \\ \lambda_{i,j} & \text{en otro caso} \end{cases} \\ \tilde{\lambda}_{i,j}^* &= \tilde{\lambda}_{i,j} - \left| \tilde{\lambda}_{i,j} \right| \frac{\sum_{j=1}^n \tilde{\lambda}_{i,j}}{\sum_{j=1}^n \left| \tilde{\lambda}_{i,j} \right|} \text{ para } i, j = 1, 2, \dots, n\end{aligned}$$

Los códigos para los tres métodos se muestran a continuación.

```
function Tp=ajuste_tasas1(T)

% ESTA FUNCION REALIZA EL AJUSTE DE LA MATRIZ
% DE TASAS HACIENDO USO DEL METODO 1.
%
% FORMA DE USARSE: Tp=ajuste_tasas1(T)
```

```

n=max(size(T));
Diag=diag(T); % Los elementos de la diagonal
% Modificamos los elementos de la diagonal
for k=1:n
    T(k,k)=0;
end

```

```

Tp=max(T,0);
SApoyo=sum(min(T,0),2);

```

```

for k=1:n
    Tp(k,k)=Diag(k)+SApoyo(k);
end

```

-----

```

function Tp=ajuste_tasas2(T)

```

```

% ESTA FUNCION REALIZA EL AJUSTE DE LA MATRIZ
% DE TASAS HACIENDO USO DEL METODO 2.

```

```

%
```

```

% FORMA DE USARSE: Tp=ajuste_tasas2(T)

```

```

n=max(size(T)); Diag=diag(T);
for k=1:n
    T(k,k)=0;
end

```

```

Tp=T;
G=abs(Diag) + sum(max(Tp,0),2); B=sum(max(-Tp,0),2);
ind=find(Tp<0); Tp(ind)=0;
Tp=Tp + diag(Diag);

```

```

% Modificamos los renglones

```

```

for k=1:n
    if(G(k)>0)
        Tp(k,:)=Tp(k,:) - B(k).*abs(Tp(k,:))./G(k);
    end
end

```

-----

```

function Tp=ajuste_tasas3(T)

```

```

% ESTA FUNCION REALIZA EL AJUSTE DE LA MATRIZ
% DE TASAS HACIENDO USO DEL METODO AJUSTE PONDERADO.

```

```

%
% FORMA DE USARSE: Tp=ajuste_tasas3(T)

n=max(size(T)); Tp=zeros(n,n);

for k=1:n
    if(T(k,k)~=0)
        valor=T(k,k); ind=find(T(k,:)<0);
        T(k,ind)=0; T(k,k)=valor;
        suma1=sum(T(k,:)); suma2=sum(abs(T(k,:)));
        Tp(k,:)=T(k,:)-abs(T(k,:))*suma1/suma2;
    end
end
end

```

La idea es poder comparar cuál de todos estos ajustes al calcular la matriz logaritmo nos regresa la mejor matriz de tasas válida, por lo que a continuación veremos un ejemplo práctico. Para ello tomemos la siguiente matriz de transición

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.2 & 0.2 & 0.2 \\ 0.1 & 0.6 & 0 & 0.3 \\ 0.3 & 0.2 & 0.5 & 0 \\ 0.1 & 0.1 & 0.2 & 0.6 \end{bmatrix} \quad \ln(\mathbf{P}) = \begin{bmatrix} -1.1628 & 0.3496 & 0.4390 & 0.3742 \\ 0.2240 & -0.5612 & -0.1358 & 0.4730 \\ 0.7234 & 0.2734 & -0.7969 & -0.2006 \\ 0.0731 & 0.0999 & 0.3636 & -0.5366 \end{bmatrix}$$

Lo que se desea saber de manera general cuál de todos los posibles ajustes nos regresan "la mejor" matriz de transición.

```

>> P=[0.4 0.2 0.2 0.2; 0.1 0.6 0 0.3;
      0.3 0.2 0.5 0; 0.1 0.1 0.2 0.6];

>> TQ=matriz_generadora_IRW(P);
>> % Salen solamente dos matrices
>> T=real(TQ(:,:,2))
T =
    -1.1628    0.3496    0.4390    0.3742
     0.2240   -0.5612   -0.1358    0.4730
     0.7234    0.2742   -0.7969   -0.2006
     0.0731    0.0999    0.3636   -0.5366

>> T_JLT=generador_JLT(P)
T_JLT =
    -0.9163    0.3054    0.3054    0.3054
     0.1277   -0.5108         0    0.3831
     0.4159    0.2773   -0.6931         0
     0.1277    0.1277    0.2554   -0.5108

```

```

>> Taj1=ajuste_tasas1(T)
Taj1 =
    -1.1628    0.3496    0.4390    0.3742
     0.2240   -0.6970         0    0.4730
     0.7234    0.2742   -0.9976         0
     0.0731    0.0999    0.3636   -0.5366

>> Taj2=ajuste_tasas2(T)
Taj2 =
    -1.1628    0.3496    0.4390    0.3742
     0.1998   -0.6218         0    0.4220
     0.6425    0.2435   -0.8860         0
     0.0731    0.0999    0.3636   -0.5366

>> Taj3=ajuste_tasas3(T)
Taj3 =
    -1.1628    0.3496    0.4390    0.3742
     0.1998   -0.6218         0    0.4220
     0.6425    0.2435   -0.8860         0
     0.0731    0.0999    0.3636   -0.5366

```

De manera inicial se puede observar que la matriz usando el método *JLT* es la que más difiere de las demás, y los otros tres métodos nos regresan matrices ajustadas similares. Por lo que es necesario poder comparar cuál de todas estas matrices representa mejor a la matriz inicial de transición. Existen varios métodos de comparación. El más fácil de proponer es calcular la distancia  $\|\mathbf{P} - e^{\tilde{\mathbf{T}}}\|_2$ , otra propuesta que se indica en Israel et. al [58] es quedarnos con la matriz de tasas que minimice la suma

$$S = \sum_{i,j} |j - i| |\lambda_{i,j}| \quad (4.22)$$

esto garantiza minimizar que haya un cambio de estado grande muy pronto. Por último, en Jafry y Schuermann [73] se propone la siguiente métrica

$$D_{\text{SVD}}(\mathbf{A}, \mathbf{B}) = M_{\text{SVD}}(\mathbf{A}) - M_{\text{SVD}}(\mathbf{B}) \quad (4.23)$$

$$\text{con } M_{\text{SVD}}(\mathbf{Q}) = \frac{1}{n} \sum_{k=1}^N \sqrt{\lambda_k (\mathbf{Q} - \mathbf{I})^T (\mathbf{Q} - \mathbf{I})}$$

donde se utiliza la descomposición SVD<sup>5</sup> y es el promedio de los valores singulares de la matriz, y que la ventaja de esta métrica es que como los valores singulares de una matriz

<sup>5</sup>En la descomposición singular SVD de una matriz  $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$ , los valores singulares de una matriz  $\mathbf{A}$  son la raíz cuadrada de los valores propios de  $\mathbf{A}^T\mathbf{A}$ , y estos están contenidos en la matriz diagonal  $\mathbf{\Sigma}$

$$\sigma_i(\mathbf{A}) = \sqrt{\lambda_i(\mathbf{A}^T\mathbf{A})}$$

siempre son positivos se puede interpretar mejor como una medida, ya que los valores propios pueden ser negativos y hacer difícil el análisis. Otra ventaja es que al sustraer la matriz identidad, uno se queda con la parte de la movilidad de la matriz, ya que la matriz identidad representa una matriz de transición estática. A continuación se hará un m-file que permita analizar fácilmente con cuál matriz nos quedaremos para analizar el problema de inclusión.

```
function Valores = analisis_matrices_tasas(P,T)

% ESTA FUNCION REGRESA LOS DISTINTOS VALORES AL
% COMPARAR LOS DISTINTOS AJUSTES. "P" ES LA MATRIZ
% DE TRANSICION Y "T" ES LA MATRIZ LOGARITMO SIN
% AJUSTAR. LA MATRIZ "Valores" EN CADA COLUMNA
% REGRESA LOS RESULTADOS POR CADA METODO DE COMPARACION
%
% FORMA DE USARSE: Valores = analisis_matrices_tasas(P,T)

n=max(size(P));

T_JLT = generador_JLT(P); Taj1=ajuste_tasas1(T);
Taj2=ajuste_tasas2(T); Taj3=ajuste_tasas3(T);

E_JLT=expm(T_JLT); E1 = expm(Taj1);
E2 = expm(Taj2); E3 = expm(Taj3);

S=salto(n);
S_JLT=sum(sum(abs(T_JLT).*S)); Saj1=sum(sum(abs(Taj1).*S));
Saj2=sum(sum(abs(Taj2).*S)); Saj3=sum(sum(abs(Taj3).*S));

M_P = M_SVD(P);
M_JLT=M_SVD(E_JLT); Maj1=M_SVD(E1); Maj2=M_SVD(E2); Maj3=M_SVD(E3);

Valores(1,1)=norm(P-E_JLT,2); Valores(1,2)=S_JLT; Valores(1,3)=M_P - M_JLT;
Valores(2,1)=norm(P-E1,2); Valores(2,2)=Saj1; Valores(2,3)=M_P - Maj1;
Valores(3,1)=norm(P-E2,2); Valores(3,2)=Saj2; Valores(3,3)=M_P - Maj2;
Valores(4,1)=norm(P-E3,2); Valores(4,2)=Saj3; Valores(4,3)=M_P - Maj3;

%-----
function M=salto(n)

M=zeros(n);
for i=1:n
    for j=1:n
        M(i,j)=abs(j-i);
    end
end
end
```

```
%-----
function prom=M_SVD(P)

I = eye(size(P));
prom=mean(svd(P-I));
```

Con este m-file será más fácil decidir con que matriz de tasas, con el ejemplo que se había analizado previamente obtendríamos los siguientes resultados

```
>> Valores = analisis_matrices_tasas(P,T)
Valores =
    0.1358    4.7295    0.0452
    0.1339    6.0240   -0.0197
    0.1170    5.7053   -0.0025
    0.1170    5.7053   -0.0025
```

Esto implica que la mejor matriz de tasas se obtiene con el ajuste del método 2 y 3 (renglón 3 y 4), ya que se minimiza la distancia y la métrica

$$\left\| \mathbf{P} - e^{\tilde{\mathbf{T}}_{aj3}} \right\|_2 \text{ y } D_{\text{SVD}}(\mathbf{P}, e^{\tilde{\mathbf{T}}_{aj3}})$$

y esto nos dará una mejor información al trabajar a tiempo continuo con una cadena de Markov. También se observa que con el ajuste *JLT* no se obtiene una buena aproximación, esto podría considerarse debido a las hipótesis que los autores proponen que provienen más de hipótesis financieras, pero que en un contexto más general no nos regresa una matriz de tasas que represente a la cadena de Markov original.

#### 4.3.4 Un caso de estudio. Avance académico estudiantil

Como se había comentado anteriormente, la aplicación recurrente sobre el problema de inclusión es el análisis del cambio de calificación crediticia de una compañía en un tiempo determinado donde se tiene una cantidad finita de estados, donde el primer estado representa la mejor calificación y el último estado representa insolvencia económica, y se puede estudiar más a fondo estos estudios en Lando [56], Israel et. al [58], Inamura [71], Jarrow et. al [57] y Jafry [72] por citar a algunos. En esta parte se trabajará con una metodología similar pero se analizará la trayectoria académica de estudiantes en el nivel medio superior. Uno de los problemas actuales que se presentan en el nivel medio superior es la alta deserción, que en algunos casos puede llegar a ser del 50%<sup>6</sup>, esta deserción es multifactorial y no puede atribuirse únicamente a las instituciones educativas. En el año 2012 la Secretaría de Educación Pública publicó la Encuesta Nacional de Deserción en la Educación Media Superior, que analiza distintos factores del porqué se da este fenómeno, citando "Los condicionantes analizados abarcan desde la organización de los sistemas educativos, el contexto social, la gestión escolar, la relación de los alumnos con los docentes, la situación familiar y

<sup>6</sup><http://www.milenio.com/cdb/doc/noticias2011/c63cbf536891f237e48836d3233261eb>

la situación individual" que en el caso del Distrito Federal y el estado de Nuevo León son las entidades donde se tiene un mayor porcentaje de deserción. Para nuestro caso en particular se estudiará la movilidad estudiantil de una de las preparatorias del Instituto de Educación Media Superior del Gobierno del Distrito Federal. El Sistema de Bachillerato del DF nace a partir del año 1997 cuando el Ing. Cuauhtémoc Cárdenas decide construir una preparatoria en la delegación Iztapalapa, en lo que antes era la Cárcel de Mujeres, esto fue debido a que los colonos demandaban utilizar como centro educativo las instalaciones de la ex Cárcel de Mujeres que se encontraban en remodelación. Fue hasta el año 2000 que se publica el Decreto de Creación del Instituto de Educación Media Superior del DF (IEMS-DF). En la actualidad el IEMS-DF cuenta con 20 planteles repartidos en la mayoría de las delegaciones políticas y aproximadamente con una matrícula estudiantil de 25,000 estudiantes. Es importante notar que las preparatorias están ubicadas en zonas donde no existía una oferta académica de este nivel, y que podría considerarse como escuelas locales, no pueden ingresar jóvenes que no vivan en un radio "cercano" a las escuelas, las colonias que estan dentro de la zona de convergencia se determinan cada año.

Una de las principales características del sistema de bachillerato del IEMS-DF comparándola con otras instituciones académicas, es que no se tiene una calificación numérica, sino que cada asignatura tiene solamente dos opciones: "Cubre" o "No Cubre". Cuando un estudiante ha cubierto una materia esto implica que tiene los conocimientos básicos adquiridos, tiene una actitud positiva hacia el trabajo académico y una serie de habilidades mínimas por asignatura. El plan de estudios actual a partir del año 2004 es el siguiente, repartida en seis semestres de trabajo.

<b>Primer semestre</b>	<b>Segundo semestre</b>	<b>Tercer semestre</b>
Matemáticas I	Matemáticas II	Matemáticas III
Física I	Física II	Química I
Lengua y Literatura I	Lengua y Literatura II	Lengua y Literatura III
Filosofía I	Filosofía II	Filosofía III
POE I	POE II	Historia I
Computación I	Computación II	Artes Plásticas I
<b>Cuarto semestre</b>	<b>Quinto semestre</b>	<b>Sexto semestre</b>
Matemáticas IV	Matemáticas V	Biología II
Química II	Biología I	Historia IV
Lengua y Literatura IV	Inglés II	Música II
Filosofía IV	Música I	Inglés III
Historia II	Optativa área humanidades	Optativa área humanidades
Artes Plásticas II	Optativa área ciencias	o area ciencias
Inglés I		Problema Eje

En total el estudiante tiene que cubrir satisfactoriamente 38 asignaturas. Cada asignatura esta compuesta por distintos objetivos temáticos, y hasta antes del año 2006 un estudiante debería tener cubiertos cada uno de estos objetivos temáticos para poder tener cubierta satisfactoriamente la asignatura. A partir del año 2008 se cambió a la evaluación compendiada, por lo que ya si cada docente considera que un estudiante tiene los elementos suficientes

para acreditar una asignatura pero le hace falta algunos objetivos, puede dar por cubierta la asignatura y recomendarle a cada estudiante el plan de trabajo académico para que mejore en los aspectos pendientes. También es importante decir que cuando un estudiante no cumple con los elementos para acreditar una asignatura, existen dos mecanismos distintos lograr este objetivo. El primer caso es que el estudiante tenga que recurrir toda la asignatura debido a que no cuenta con los elementos mínimos y en el otro caso se pueden trabajar los elementos pendientes en los módulos de recuperación o periodos de recuperación. Un estudiante de primer ingreso cuenta inicialmente con dos periodos de evaluación compendiado, al final del semestre o en el periodo intersemestral. Si al final de estos periodos un docente recomienda módulos de recuperación, el estudiante actualmente cuenta con dos periodos durante el semestre normal para inscribirse en los periodos de recuperación y poder cubrir la asignatura correspondiente. Antes del 2008 se llevaban tres periodos durante el semestre para cubrir una asignatura en esta modalidad de módulos de recuperación, al final se lleva a cabo la evaluación de las materias inscritas en curso regular y en el periodo intersemestral. La asignatura de "Problema Eje" tiene dos propósitos centrales, el primero es que el estudiante desarrolle un trabajo de algún tema de interés y que aplique los conocimientos adquiridos durante su estancia en el bachillerato, y al final de la exposición se le asignará una evaluación numérica la cuál aparecerá en su certificado de bachillerato.

El ingreso al IEMS-DF es mediante sorteo ante notario público, no existe un examen de admisión como lo hay para ingresar a otros sistemas educativos, examen administrado por la Comisión Metropolitana de Instituciones Públicas de Educación Media Superior (COMIPEMS). Esto ha implicado que algunos estudiantes no confíen en la validez de los resultados académicos, y por lo tanto abandonen la escuela en el primer año de estudios. Otros estudiantes han externado que se inscriben al IEMS-DF de manera paralela a otras instituciones públicas, si al realizar su examen de admisión en el COMIPEMS obtienen un resultado favorable en algunas de sus opciones se quedan en esta escuela y ya no concluyen su inscripción formal al IEMS. En el otro caso también si no se quedan en su opción deseada y el COMIPEMS les asigna una institución que no satisface sus expectativas académicas, entonces trabajan un año en el IEMS y después abandonan la escuela, por lo que se ha observado una alta deserción durante el primer año de estudios.

## Metodología

En nuestro caso se analizará el avance académico estudiantil del plantel "Belisario Domínguez" de la generación 2004 a la generación 2009, ya que en general la mayoría de los estudiantes de la generación 2009 ingresaron en agosto del 2008 por lo que hasta estos momentos o ya concluyeron sus estudios o ya abandonaron la institución, existen algunos que todavía están inscritos como estudiantes independientes, y que todavía su ciclo escolar no ha concluido. Inicialmente se estimará una matriz de transición discreta donde el proceso es la cantidad de materias cubiertas hasta la  $n$ -ésima etapa de evaluación. Se considera que un semestre está compuesto por 5 etapas de evaluación<sup>7</sup>, las cuáles las tres primeras representan los periodos

---

<sup>7</sup>A partir del año 2006 solamente existen cuatro periodos, dos para los periodos de recuperación, uno para el final del semestre y por último el del periodo intersemestral. Aunque en el análisis que haremos se tomarán los cinco periodos, debido a que así fue recibida la información por parte de la Dirección de Informática y

o módulos de evaluación, el cuarto representa el final del semestre y el quinto periodo es cuando finaliza el periodo intersemestral. Algo importante de subrayar es que las etapas de evaluación se pueden considerar equiespaciadas, y esto no alterará el análisis de fondo. Para ello se tomará el espacio de estados de la forma  $\mathbb{E} = \{0, 1, 2, \dots, 38, A\}$  donde el estado final  $A$  representa que un estudiante abandonó la escuela y no concluyó sus estudios y cada estado  $j \in \{0, 1, \dots, 38\}$  representa la cantidad de asignaturas cubiertas. Entonces el proceso  $\{X_n\}_{n \in \mathbb{N}}$  representa el avance académico de un estudiante y tendrá 40 estados. Es importante asumir que se trata de una cadena de Markov debido a que no importa toda la historia que he tenido un estudiante en el bachillerato, importa únicamente la cantidad de materias que tiene cubierto hasta el último periodo de evaluación correspondiente.

$$P [X_{n+1} = k + \# \mid X_n = k, (X_0, X_1, \dots, X_{n-1}) \in \mathcal{B}] = P [X_{n+1} = k + \# \mid X_n = k]$$

Para nuestro análisis el último estado  $A$  representa conjuntamente dos opciones, la primera fue aquel estudiante que dejó de inscribirse a los cursos regulares y no formalizó su baja definitiva, y la otra opción es que el IEMS los dio de baja debido a un bajo aprovechamiento escolar durante dos semestres consecutivos, el cual consiste en no aprobar más de la mitad asignaturas durante dos semestres consecutivos<sup>8</sup>. Con esto entonces los estados 38 y  $A$  serán absorbentes y los estados  $0, 1, 2, \dots, 37$  serán estados transitorios.

Para poder construir la matriz de movilidad por generación se le solicitó al IEMS una matriz de datos que especificará la cantidad de materias cubiertas en cada periodo de trabajo, como se muestra en el siguiente diagrama

			2004-2005 A					2004-2005 B				
	Estudiante	generacion	PERIODO 1	PERIODO 2	PERIODO 3	SEMESTRAL	INTERSEMESTRAL	PERIODO 1	PERIODO 2	PERIODO 3		
4	1	2004	0	0	0	6	0	0	0	0		
5	2	2004	0	0	0	0	0	0	0	0		
6	3	2004	0	0	0	5	1	0	0	0		
7	4	2004	0	0	0	0	1	1	0	0		
8	5	2004	0	0	0	1	1	0	0	0		
9	6	2004	0	0	0	0	1	0	0	0		
10	7	2004	0	0	0	2	2	1	0	0		
11	8	2004	0	0	0	6	0	0	0	0		
12	9	2004	0	0	0	3	2	0	0	0		
13	10	2004	0	0	0	3	1	0	1	0		
14	11	2004	0	0	0	0	0	0	0	0		
15	12	2004	0	0	0	3	3	0	0	0		
16	13	2004	0	0	0	2	2	1	0	1		
17	14	2004	0	0	0	0	1	0	0	0		
18	15	2004	0	0	0	3	2	0	0	0		
19	16	2004	0	0	0	2	3	0	0	0		
20	17	2004	0	0	0	1	0	0	0	0		
21	18	2004	0	0	0	4	2	0	0	0		
22	19	2004	0	0	0	0	1	0	0	0		
23	20	2004	0	0	0	0	5	0	1	0		
24	21	2004	0	0	0	4	1	1	0	0		

y después se modificó esta matriz de datos original para obtener la cantidad de materias

Telecomunicaciones del IEMS, donde no hubo una distinción entre los estudiantes que ingresaron antes del 2006 y los que ingresaron después de este año.

<sup>8</sup>Antes del 2006 un estudiante tenía la oportunidad de terminar sus estudios en el tiempo que fuera necesario, pero después se modificaron las reglas de inscripción en el cuál el estudiante solo tiene cuatro años y medio para concluir su bachillerato teniendo que cubrir más de la mitad de las materias cubiertas cada semestre.

acumuladas cubiertas hasta cada periodo de evaluación, como se muestra a continuación

Estudiante	generación	2004-2005A					2004-2005B			
		PERIODO 1	PERIODO 2	PERIODO 3	SEMESTRAL	INTERSEMESTRAL	PERIODO 1	PERIODO 2	PERIODO 3	SEI
1	2004	0	0	0	6	6	6	6	6	
2	2004	0	0	0	0	0	0	0	0	
3	2004	0	0	0	5	6	6	6	6	
4	2004	0	0	0	0	1	2	2	2	
5	2004	0	0	0	1	2	2	2	2	
6	2004	0	0	0	0	1	1	1	1	
7	2004	0	0	0	2	4	5	5	5	
8	2004	0	0	0	6	6	6	6	6	
9	2004	0	0	0	3	5	5	5	5	
10	2004	0	0	0	3	4	4	5	5	
11	2004	0	0	0	0	0	0	0	0	
12	2004	0	0	0	3	6	6	6	6	
13	2004	0	0	0	2	4	5	5	6	
14	2004	0	0	0	0	1	1	1	1	
15	2004	0	0	0	3	5	5	5	5	
16	2004	0	0	0	2	5	5	5	5	
17	2004	0	0	0	1	1	1	1	1	
18	2004	0	0	0	4	6	6	6	6	
19	2004	0	0	0	0	1	1	1	1	
20	2004	0	0	0	0	5	5	6	6	
21	2004	0	0	0	4	5	6	6	6	

El avance académico de cada estudiante se puede pensar como una realización del proceso  $\{X_n\}_{n \in \mathbb{N}}$ , y se asume que todos los estudiantes cuentan con las mismas condiciones académicas y sociales, aunque esto no es cierto necesariamente. Aquí es necesario recalcar que con los datos obtenidos se tuvo que estimar el momento en que algún estudiante llegó a abandonar sus estudios, en cualquiera de las dos opciones que se especificaron anteriormente. Esto se puede observar cuando el número de materias cubiertas ya se queda fija por un largo periodo de tiempo. Este análisis se tuvo que realizar estudiante por estudiante. Para poder estimar la matriz de transición  $\hat{P}_{i,j}$  se hará uso de la estimación máximo verosímil que se desarrolla en Anderson y Goodman [70] y está dada por

$$\hat{P}_{i,j} = \frac{N_{i,j}}{N_i^*} = \frac{\sum_{n=1}^T N_{i,j}(n, n+1)}{\sum_{j \in \mathbb{E}} \sum_{n=1}^T N_{i,j}(n, n+1)} \tag{4.24}$$

donde  $N_{i,j}(n, n+1)$  representa el número de veces que el proceso paso del estado  $i$  al estado  $j$  de la etapa  $n$  a la etapa  $n+1$ , y  $N_i^*$  es el número total de veces que el proceso se encontró en el estado  $i$ . En este caso se asumirá que el proceso es homogéneo, porque en la práctica se tendrían dos probabilidades distintas en tiempos distintos, por ejemplo,  $P[X_1 = 0 | X_0 = 0] \neq P[X_6 = 0 | X_5 = 0]$ , y con ello se tiene el inconveniente de que en general el proceso es no homogéneo, y que entonces por cuestiones de modelación se asumirá que el proceso de avance académico es homogéneo. Otra cuestión importante es que con la información de un solo estudiante no se podría estimar todo el proceso, por lo que cada uno de los estudiantes conformará una realización del proceso, y con ello se podrá construir la matriz de transición por generación.

```

function Pglobal=matriz_global(MatDatos)

% ESTA FUNCION REGRESA LA MATRIZ GLOBAL DE TRANSICION
% PARA LA MATRIZ DE DATOS "MatDatos" DONDE CADA RENGLON
% DE LA MATRIZ REPRESENTA UNA REALIZACION DEL PROCESO.
%
% Pglobal=matriz_global(MatDatos)

% Verificamos la dimension de la matriz
[Nest,m]=size(MatDatos);

% Para ir acumulando los datos
Nglobal=zeros(40,40);

for k=1:Nest
    Temp=estimacion_matriz(MatDatos(k,:));
    Nglobal = Nglobal + Temp;
end

% Dividimos al final entre el total de realizaciones
Suma=sum(Nglobal,2);
Pglobal=diag(1./Suma)*Nglobal;
spy(Pglobal)

%-----
%-----

function Nij=estimacion_matriz(datos)

% ESTA FUNCION REGRESA LA ESTIMACION DE LA
% MATRIZ DE TRANSICION PARA EL VECTOR "datos"
%
% FORMA DE USARSE: Pest=estimacion_matriz(datos)

% Es una matriz con 40 estados
Nij=zeros(40,40);
T=length(datos);

% EL ESTADO 39 ES EL ESTADO "A"
for k=1:(T-1) % AQUI RECORREMOS TODO EL VECTOR "DATOS"
    for i=1:40 % AQUI ANALIZAREMOS EL ESTADO "i"
        for j=1:40 % AQUI ANALIZAREMOS EL ESTADO "j"
            if( datos(k)==(i-1) && datos(k+1)==(j-1) )
                Nij(i,j)=Nij(i,j)+1; % CAMBIO DE ESTADO i->j
            end
        end
    end
end

```

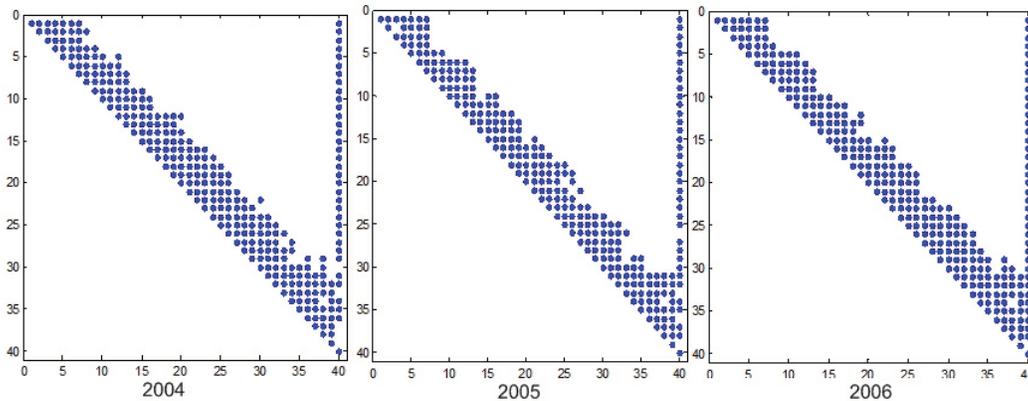
```

end
end
end
    
```

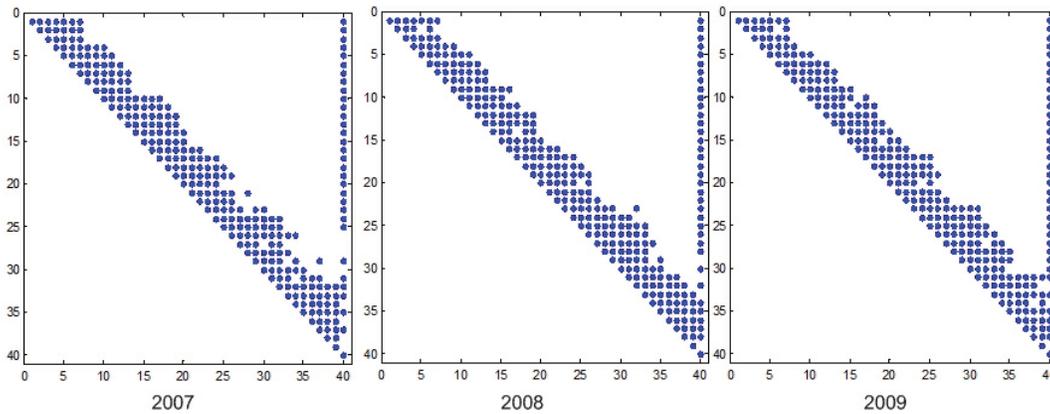
La matrícula de estudiantes por generación ronda entre los 350 estudiantes, con los datos obtenidos del IEMS-DF se tienen los siguientes datos por generación<sup>9</sup>

Generación	2004	2005	2006	2007	2008	2009
No. estudiantes	358	351	353	349	346	330

y analizando cada generación se obtienen las siguientes estructuras matriciales para las generaciones 2004, 2005 y 2006



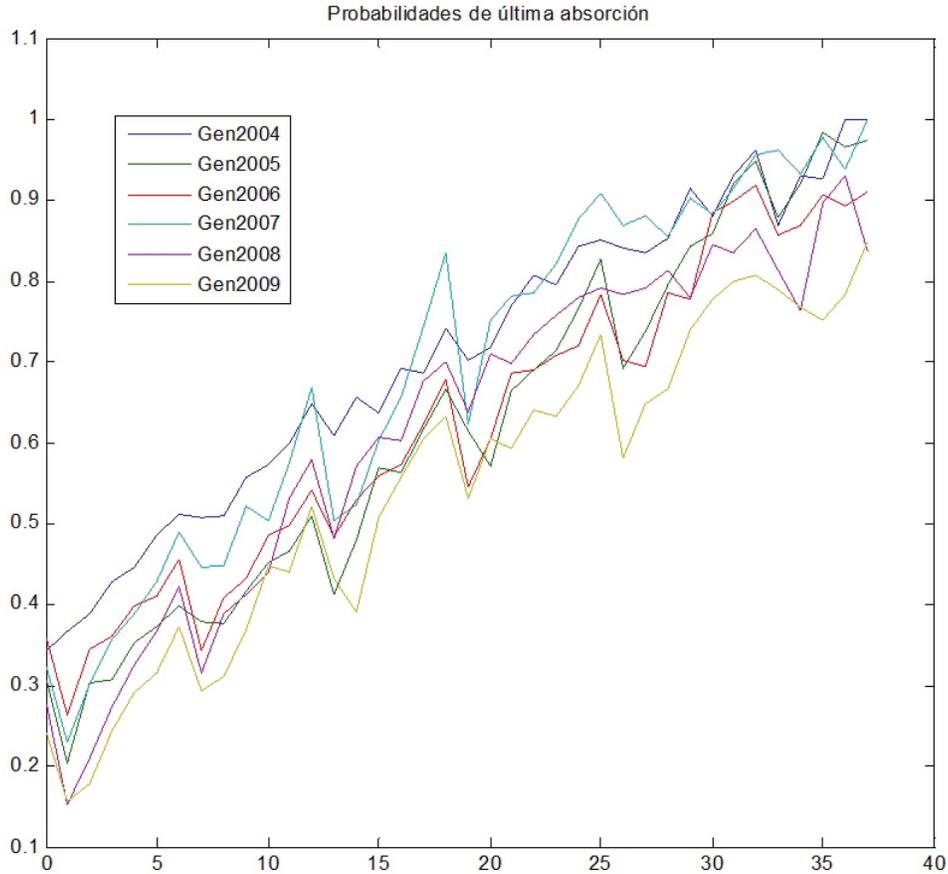
Por último, para las generaciones 2007, 2008 y 2009 tendrían la siguiente estructura



Lo que representa la última columna de cada matriz es la probabilidad de que un estudiante abandone sus estudios (por deserción o por baja), y que el dato que más sorprende

<sup>9</sup>Para las generaciones 2008 y 2009 se nos proporcionó información de 359 y 363 estudiantes respectivamente, pero el IEMS instrumentó una modalidad conocida como "estudiante independiente", que consta de que un estudiante puede estar avanzando a un ritmo más lento para concluir sus estudios, solamente deben de tener 20 asignaturas cubiertas, por lo que los estudiantes que se eliminaron de la matriz de tasas siguen estudiando actualmente.

inicialmente es que a pesar de que un estudiante pueda tener más de 30 materias cubiertas por alguna razón abandone la escuela, ya que por lo menos le invirtió dos años y medio, y que la razón más probable para no concluir sus estudios sea por cuestiones económicas. A continuación se graficarán las probabilidades de última absorción (3.34) para las seis generaciones

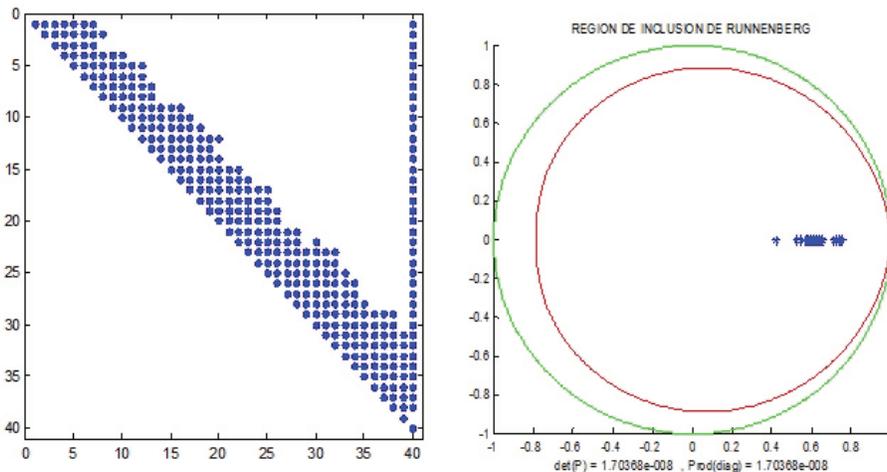


El dato más revelador y preocupante nos la dice  $F_{0,38}$ , que representa la probabilidad de que eventualmente un estudiante que ingresa a la preparatoria "Belisario Domínguez" concluya sus estudios. Se muestran estas probabilidades de última absorción

	2004	2005	2006	2007	2008	2009
$F_{0,38}$	34.35%	30.76%	35.97%	32.36%	27.74%	24.17%
Obs.	34.63%	31.05%	35.57%	32.37%	27.74%	23.93%

con lo que podemos observar que se tiene un buen modelo matemático para representar el avance académico general de un estudiante, aunque en general los datos de deserción escolar a nivel nacional son altos, en nuestro caso pues estamos por debajo de la media, y se tiene que tomar en cuenta el entorno socioeconómico de los estudiantes. Lo que se desea saber ahora es si cada una de estas matrices puede provenir de algún proceso de Markov, y para ello usaremos las herramientas expuestas anteriormente. En nuestro caso tenemos dos estados absorbentes, por lo que el valor propio  $\lambda = 1$  tiene multiplicidad 2, lo que implica para nosotros solamente poder estimar la matriz logaritmo usando la serie de potencias (4.17) y (4.18). Lo que se hará es obtener una matriz de transición promedio de todas las generaciones,

y con esta nueva matriz construir una matriz de tasas generadora.



Todos los valores propios son reales y positivos, por lo que esto al menos garantiza la existencia de la matriz logaritmo. El problema con esta matriz es que no existirá una generadora exacta debido a que no se cumple la hipótesis de que si  $i \rightarrow j$  entonces  $P_{i,j} > 0$ , como por ejemplo un estudiante puede concluir satisfactoriamente su bachillerato, sabemos que  $P_{0,38}^{(n)} > 0$  para alguna  $n \in \mathbb{N}$  pero en la matriz de transición estimada  $P_{0,38} = 0$ . Haciendo el análisis en Mat/Oct

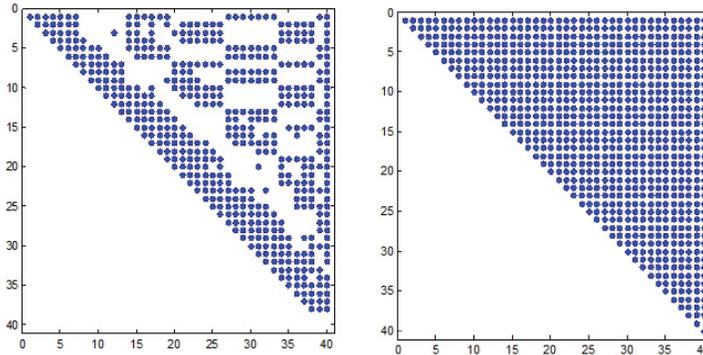
```
>> T=logaritmo_matriz_serie(Pgeneral,500);
RADIO ESPECTRAL MENOR QUE 1, DADO POR 0.576956

>> % La matriz Pgeneral fue el promedio de las
>> % seis matrices generacionales

>> Valores = analisis_matrices_tasas(Pgeneral,T)
Valores =
    0.1096    77.0070     0.0043
    0.0276    68.5111    -0.0061
    0.0238    67.5601    -0.0023
    0.0238    67.5601    -0.0023
```

Esto nos demuestra que se usará el método 3 (que esta representado por el último renglón de la matriz "Valores", se minimizan las  $\|\mathbf{P} - e^{\mathbf{T}}\|$ , y las estadísticas (4.22) y (4.23) ) para realizar el ajuste correspondiente a la matriz logaritmo original. Aunque se debe tener mucho cuidado ya que al realizar el ajuste correspondiente quedarán algunas entradas positivas con probabilidades muy pequeñas que en la matriz de transición original no aparecían, como se

muestra a continuación

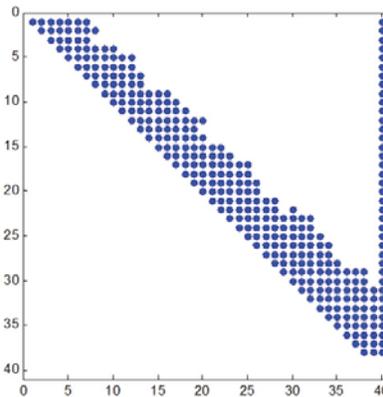


La matriz logaritmo ajustada (izquierda) y la matriz exponencial  $e^{\mathbf{T}_{aj3}}$  (derecha). Se observa la perdida general de la estructura matricial original.

Los elementos que aparecen en la matriz logaritmo fuera de la " diagonal" ahora permitirán probabilidades que en el modelo original no existían, como por ejemplo que un estudiante pueda concluir sus estudios en el primer periodo de evaluación

$$\begin{array}{ll} \text{Original} & P[X_1 = 38 \mid X_0 = 0] = 0 \\ \text{Modificada} & P[X_1 = 38 \mid X_0 = 0] = 2.77818 \times 10^{-8} \end{array}$$

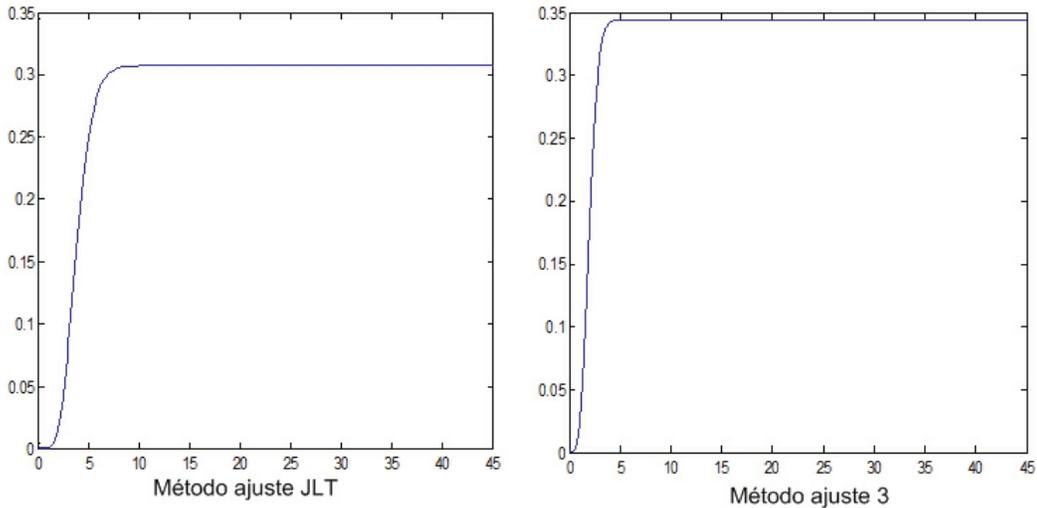
pues aunque es una probabilidad muy pequeña no debería de existir, por lo que muestra las dificultades teóricas de querer analizar un proceso discreto e incluirlo en un proceso continuo, y que se deben de cumplir muchas hipótesis teóricas. En este caso, el ajuste que más convendría utilizar es el del método *JLT*, ya que preservaría la estructura matricial original.



Ajuste metodo *JLT*

Se usarán estos dos métodos para poder obtener las probabilidades de última absorción y las probabilidades de transición  $P_{0,38}(t)$  en el intervalo  $[0, 45]$  (que sería obtener las prob-

abilidades de que un estudiante egrese en sus cuatro años y medio años reglamentarios).

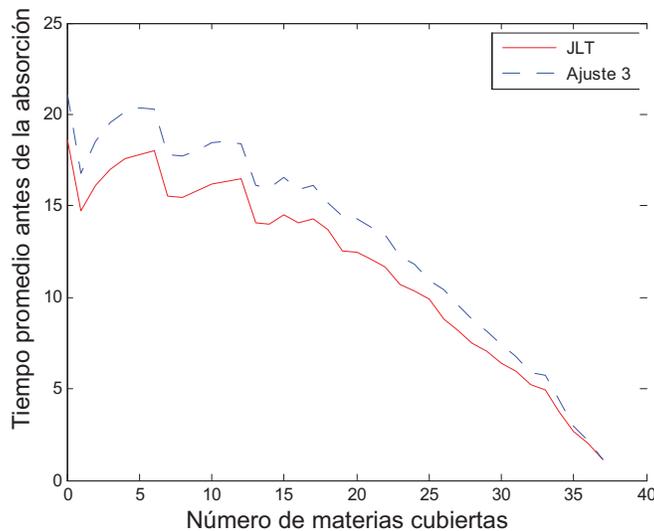


$(t, P_{0,38}(t))$  con  $t \in [0, 45]$

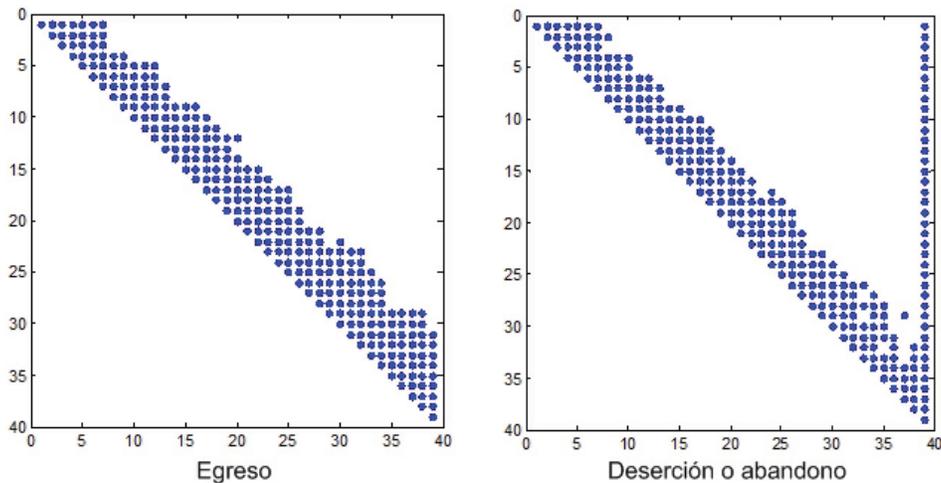
Ambas son probabilidades parecidas, y nos dan una idea de que si se toma el promedio de las seis generaciones, el egreso general sería de alrededor del 30%, lo cual indica que existen otros factores externos a la institución por los cuáles los estudiantes están abandonando sus estudios, que existen factores que la Institución debe considerar para poder revertir este proceso de deserción. Las probabilidades de última absorción para ambos métodos se muestran a continuación

	Método <i>JLT</i>	Método ajuste 3
$F_{0,38}$	30.70%	34.47%

Otro dato importante que se puede inferir de este modelo es el tiempo promedio de que un estudiante dependiendo del número de materias cubiertas egrese o abandone la escuela. Para ello se graficará el vector de  $\mathbf{M}^* = \mathbf{M}\mathbf{e}'$  con  $\mathbf{M}$  la matriz fundamental.

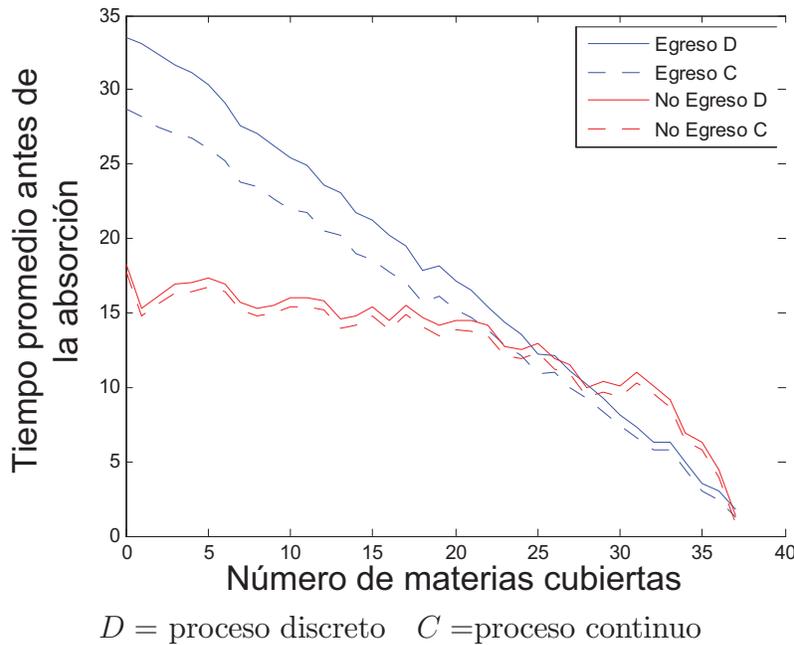


La evidencia empírica observada en el plantel "Belisario Domínguez" hasta el año 2013 es que muchos estudiantes desertan en el primer año de la escuela, ya sea porque han decidido cambiarse a otra institución pública o porque se encuentran bajo ciertas circunstancias familiares o personales que les impiden asistir. La gráfica nos indica que es más probable que un estudiante que tiene 0 asignaturas cubiertas abandone la escuela y que lo hará en el tiempo promedio estimado de dos años (18.58 unidades de tiempo con el método JLT y 21.11 unidades con el método del ajuste 3), y cuando un estudiante tiene un mayor número de asignaturas cubiertas es menos probable que abandone la escuela y por consiguiente concluya satisfactoriamente sus estudios, como es el caso de quienes tienen 30 asignaturas cubiertas y que el tiempo promedio en el que egresarán es de aproximadamente un semestre y un periodo de recuperación (6.38 unidades de tiempo con JLT y 7.39 unidades de tiempo con el ajuste 3). Anteriormente se analizaron conjuntamente los dos tipos de estudiantes, los que egresaron y no egresaron por generación, pero ahora se desea analizar por separado cada tipo de estudiante. En este caso, al juntar a todos los estudiantes egresados y por el otro a los que abandonaron, se asume que las condiciones son iguales, que es un proceso homogéneo en general. Para ello se construyeron las matrices de transición por tipo de estudiante (ahora cada matriz solamente tiene 39 estados), por lo que se tuvieron unos pequeños cambios a los códigos de estimación de la matriz. Analizando estos dos casos por separado se tienen las siguientes estructuras de las matrices de transición



Para la matriz de los estudiantes egresables si es posible calcular la matriz logaritmo usando el método IRW debido a que todos los valores propios son distintos. Para el caso de la matriz de los estudiantes que abandonan la escuela no es posible calcular la matriz logaritmo con el método IRW debido a que el valor propio  $\lambda = 0.63$  tiene multiplicidad algebraica 2, por lo que se tiene que utilizar el método de serie de potencias. Con ello, se podrán calcular los tiempos medios de absorción para el caso discreto y continuo de cada proceso, y nos dará la información más precisa de cuanto tiempo un estudiante terminará

sus estudios o abandonará la escuela.



En nuestro caso, para los estudiantes que si egresan de la preparatoria, en promedio se tardan 3 años y medio en concluir sus estudios y que en el caso continuo subestima estos tiempos diciendo que un estudiante concluye sus estudios en aproximadamente en 29 periodos aproximadamente (tres años sin periodo intersemestral). Esta forma de comparar los tiempos promedio de egreso nos diria que el problema de inclusión para el caso de estudiantes que egresan no es tan válido, que el caso discreto original nos regresa información más precisa y que en el caso de los estudiantes que no egresan los dos modelos regresan información muy similar: de que un estudiante que no egresa abandonará sus estudios en aproximadamente cuatro semestres aproximadamente, aunque con los datos observados muchos estudiantes abanadonan la escuela durante el primer año de estancia, debido a que buscan entrar a otra institución educativa. Pero el tiempo aumenta considerablemente ya que hay otros estudiantes que siguen estudiando varios semestres más, tienen un avance académico considerable pero al final abandonan debido a que tienen que ingresar al mercado laboral o tienen problemas personales graves que les impiden concluir sus estudios, y por eso el tiempo promedio para abandonar los estudios es más o menos constante, no importando si no tienen ninguna materia cubierta o tengan veinte materias cubiertas.

## 4.4 Conclusiones

A lo largo de este trabajo se pudieron revisar los distintos métodos que hay para el cálculo de la matriz exponencial, mostrando que el método de Páde es el mejor método que hay, Stewart [13] y Higham [8] son excelentes referencias sobre el tema. Al revisar las redes estocásticas, se observó que el espacio de estados crece considerablemente al juntar varios

procesos en paralelo, por lo que los métodos numéricos se convierten en la única herramienta posible para su estudio, y que en un trabajo futuro se puede estudiar más a fondo la matriz de tasas usando técnicas sparse, ya que la matriz resultante contiene una gran cantidad de ceros. La teoría de colas resulta un tema muy importante, donde los procesos de Markov juegan un papel fundamental. En este trabajo solo se mostró una pequeña parte de las distintas configuraciones que hay, y que al juntarlos con las redes estocásticas se pueden trabajar una mayor variedad de sistemas, y esto puede servir de trabajo para el estudio de colas reales y la optimización de tiempos de espera. En el último caso, al trabajar con el problema de inclusión, el problema central es poder encontrar la matriz logaritmo, en los métodos propuestos usando series de potencias siempre existirá el riesgo de generar errores numéricos, por lo que una línea de trabajo a futuro es poder encontrar de manera exacta esta matriz. En el artículo de Cardoso [76] propone un método directo para el cálculo de la matriz logaritmo para una cierta clase de funciones, pero tiene el inconveniente de que usa el polinomio mínimo de la matriz, y este polinomio solamente puede ser obtenido usando software de cálculos simbólicos como Maple o Mathematica, un trabajo a futuro podría ser el de analizar las propiedades numéricas del método propuesto en Cardoso [76]. Otro método para calcular la matriz logaritmo viene en Higham [8] es el de la potencia y escala inversa,  $\log(\mathbf{A}) = 2^n \log\left(\left(\mathbf{A}^{1/2}\right)^n\right)$ , y que en este caso es necesario el estudio de como calcular la raíz cuadrada de una matriz, por lo que se muestra la gama de aplicaciones que puede llegar a tener las funciones de matrices. Por último, al trabajar con el problema de inclusión y la movilidad estudiantil, esto puede servir a futuro para estudiar el egreso y la deserción de otros sistemas educativos a nivel medio superior, con el estudio realizado se observó que la deserción académica conlleva a problemas graves en todos los sistemas educativos, donde cada subsistema educativo tiene que revisar más a fondo el avance académico de sus estudiantes, si las reglas generales de control escolar están dando los resultados deseados y como mejorar los procesos internos para evitar la deserción. Un trabajo a futuro consistirá en analizar más factores socioeconómicos y poder incorporarlos a un modelo más general.

# Bibliografía

- [1] Meyer C. (2001) "Matrix Analysis And Applied Linear Algebra", Society for Industrial and Applied Mathematics
- [2] Kolmogorov A. (1956). "Foundations Of The Theory Of Probability", traducción al inglés Nathan Morrison, Chelsea Publishing Company
- [3] García A. M. (2005). "Introducción a la Teoria de la Probabilidad. Primer curso", Fondo de Cultura Económica
- [4] García A. M. (2005). "Introducción a la Teoria de la Probabilidad. Segundo curso", Fondo de Cultura Económica
- [5] Hernández A. F. (2003). "Cálculo de Probabilidades". Aportaciones Matemáticas # 25, Sociedad Matemática Mexicana
- [6] Harris B. (1966). "Theory Of Probability", Addison-Wesley Series In Statistics
- [7] DeGroot M. H. (1988). "Probabilidad y Estadística, Addison-Wesley Iberoamericana
- [8] Higham J. N. (2008). "Functions of Matrices: Theory and Computation", Society for Industrial & Applied Mathematics, U.S.
- [9] Higham J. D., Griffiths F. D. (2010). "Numerical Methods For Ordinary Differential Equations. Initial Value Problems", Springer Undergraduate Mathematics Series
- [10] Highman J. M. (2002). "Accuracy and Stability of Numerical Algorithms. Second Edition", Society for Industrial & Applied Mathematics, U.S.
- [11] Grimmett G., Stirzaker D. (2001). "Probability And Random Processes. Third Edition", Oxford University Press
- [12] Butcher J. C. (2003) "Numerical Methods For Ordinary Differential Equations", John Wiley And Sons
- [13] Stewart J. W. (1994). "Introduction To The Numerical Solution of Markov Chains", Princeton University Press, U.S.
- [14] Davis, P. J. (1973). "The Mathematics Of Matrices. A First Book Of Matrix Theory And Linear Algebra. Second Edition", John Wiley & Sons

- [15] Klenke A. (2008). "Probability Theory. A Comprehensive Course", Universitext, Springer
- [16] Matsumoto M., Nishimura T. (1998). "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator", ACM Transactions on Modeling and Computer Simulation, Volume 8 Issue 1, Jan. 1998. pp 3-30
- [17] Cullen C. G. (1972). "Matrices And Linear Transformations. Second Edition", Dover Publications Inc.
- [18] Burden R. L. (1993). "Numerical Analysis. Fifth Edition", PWS Publishing Company
- [19] Karlin S., Taylor H. M. (1975). "A First Course In Stochastic Processes. Second Edition", Academic Press
- [20] Feller W. (1968). "An Introduction To Probability Theory and its Applications. Volumen 1. Third Edition", John Wiley & Sons
- [21] Medhi J. (2003). "Stochastic Models In Queueing Theory. Second Edition", Academic Press
- [22] Resnick S. L. (1992). "Adventures in Stochastic Processes", Birkhauser Boston
- [23] Norris J. R. (1997). "Markov Chains", Cambridge University Press
- [24] Xie A. y Beerel P. (1998). "Efficient State Classification of Finite-State Markov Chains", In DAC. páginas 605-610
- [25] Wilkinson J. H. (1965). "The Algebraic Eigenvalue Problem", Oxford University Press
- [26] Billingsley P. (1995). "Probability And Measure. Third Edition", Wiley Series in Probability and Mathematical Statistics
- [27] Denardo E. V. (1977). "Periods of Connected Networks and Powers of Nonnegative Matrices", Mathematics of Operations Research 2:20-24
- [28] Iturrarán V. U. (2001). "The Synchronous Boltzmann Machine for learning and hard-combinatoric search", Ph. D. thesis, Imperial College
- [29] Meyer C. D. (1980). "The condition of a finite Markov chain and perturbation bounds for the limiting probabilities", SIAM Journal of Algebraic and Discrete Mathematics, 1 (3): 273-283
- [30] Kemeny J. Snell L. (1976). "Finite Markov Chains", Springer-Verlag
- [31] Bremaud P. (1999). "Markov Chains: Gibbs Fields, MonteCarlo Simulation and Queues", Springer-Verlag
- [32] Karlin S., Taylor H. (1981). "A Second Course In Stochastic Processes. Second Edition", Academic Press

- [33] Bhat. U. N., Miller G. K. (2002). "Elements Of Applied Stochastic Processes. Third Edition", Wiley Series In Probability And Statistics
- [34] Kulkarni V. G. (1995) "Modeling And Analysis Of Stochastic Systems", Texts in Statistical Science. Chapman & Hall / CRC
- [35] Cinlar E. (1975). "Introduction to Stochastic Processes", Prentice Hall
- [36] Kijima M. (1997). "Markov Processes For Stochastic Modelling", Chapman & Hall
- [37] Stewart W. J. (2009). "Probability, Markov Chains, Queues, And Smulation. The Mathematical Basis Of Performance Modeling", Princeton University Press
- [38] Gross D., Harris C. M. (1985). "Fundamentals of Queueing Theory. Second Edition", John Wiley & Sons
- [39] Grassmann, W. K., Taksar M. I., Heyman D. P. (1985). "Regenerative analysis and steady-state distributions for Markov chains", Operation Research 33. 1107-1116
- [40] Gantmacher F. R (1959). "Matrix Theory. Vol I y II", Chelsea Publishing Company USA.
- [41] Rudin W. (1974). "Real and Complex Analysis", Second Edition, Mc. Graw Hill
- [42] Wilkinson J. H. (1959). "The evaluation of the zeros of ill-conditioned polynomials". Part I. Numerische Mathematik 1:150-166
- [43] Golub G. H., Wilkinson J. H. (1976). "Ill-Conditioned Eigensystems and the Computation of the Jordan Canonical Form", SIAM Review, Vol. 18., No. 4, pp 578-619, Society for Industrial and Applied Mathematics
- [44] Plateau B., Stewart W. J. (1997). "Stochastic Automata Networks", Computational Probability, pp. 113-152, Kluwer Academic Press
- [45] Horn R., Johnson C. (1991). "Topics In Matrix Analysis", Cambridge University Press
- [46] Broxson B. J. (2006). "The Kronecker Product", UNF Digital Commons, University of North Florida
- [47] Van Loan, C. F. (2000). "The ubiquitous Kronecker product", Journal of Computational and Applied Mathematics 123, págs 85-100
- [48] Moler C., Van Loan C. (2003). "Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later", SIAM Review, Vol. 45, No. 1, pp. 3-000
- [49] Strang G. (2006). "Linear Algebra and it's Applications. Fourth edition", Brooks Cole Publications
- [50] Silverman R. A. (1973). "Complex Analysis with Applications", Courier Dover Publications

- [51] Hunter J. K. (2012). "An Introduction to Real Analysis", University of California
- [52] Isaacson D., Madsen R. (1976). "Markov Chains. Theory And Applications", Wiley series in probability and mathematical statistics, John Wiley & Sons
- [53] Casella G., Robert C. (1999). "Monte Carlo Statistical Methods", Springer-Verlag New York, Inc.
- [54] Chung K. L. (1960). "Markov Chains with Stationary Transition Probabilities", Springer-Verlag
- [55] Gupta N. (2010). "Performance analysis using some queueing models", Ph. D. thesis, Jaypee Institute of Information Technology University
- [56] Lando D. (1999). "Some elements of Rating-Based Credit Risk Modeling", Advanced Fixed-Income Valuation Tools, Wiley (draft)
- [57] Jarrow R., Lando D., Turnbull S. (1997). "A Markov Model for the Term Structure of Credit Risk Spreads", *The Review Of Financial Studies*, Vol. 10, No. 2, pp. 481-523
- [58] Israel R., Rosenthal J., Wei J. (2001). "Finding generators for Markov Chains via empirical transitions matrices, with applications to credit ratings", *Mathematical Finance*, Vol. 11, No. 2, pp. 245-265
- [59] Mogens B., Sorensen M. (2009). "Efficient estimation of transitions rates between credit ratings from observations at discrete time points", *Quantitative Methods in Finance*, 9, pp. 147-160
- [60] Singer B., Spilerman S. (1976). "The Representation of Social Processes by Markov models", *The American Journal of Sociology*, Vol. 82, No. 1, pp. 1-54
- [61] Rolski T., Schmidli H., Schmidt V., Teugels J. L. (1999). "Stochastic Processes for Insurance and Finance", *Wiley Series in Probability and Statistics*
- [62] Cuthbert J. R. (1972). "On uniqueness of the logarithm for Markov semi-groups", *Journal London Mathematical Society* 4, 623-630
- [63] Cuthbert J. R. (1973). "The Logarithm Function for Finite-State Markov Semi-Groups", *J. London Math. Soc.* (2), 6 (1973), pp 524-532
- [64] Kingman J. F. C. (1962). "The imbedding problem for finite Markov chains", *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 1962, Volume 1, Issue 1, pp 14-24
- [65] Davies E. B. (2010). "Embeddable Markov Matrices", *Electronic Journal of Probability*, Vol. 15, Paper No. 47, pp 1474-1486
- [66] Kreinin A., Sidelnikova M. (2001). "Regularization algorithms for transition matrices", *Algo Research Quarterly*, Volume 4, pp 23-40

- [67] Runnenberg J. Th. (1962). "On Elfving's problem of Imbedding a time-discrete Markov chain in a continuous time one for finitely many states". Proceedings, Koninklijke Nederlandse Akademie van Wetenschappen, series A, Mathematical Sciences, 65(5), pp 536-541
- [68] Cheney W., Kincaid D. (2008). "Numerical Mathematics and Computing. Sixth edition", Thomson Brooks/Cole
- [69] Taha A. M. (2007). "Operations Research. An Introduction. Eight Edition", Pearson Prentice Hall
- [70] Anderson T. W., Goodman L. A. (1957). "Statistical Inference about Markov Chains", Annals of Mathematical Statistics, Volume 28, N. 1, pp 89-110
- [71] Inamura Y. (2006). "Estimating Continuous Time Transition Matrices From Discretely Observed Data", Bank of Japan Working Paper Series, No. 6 - E07
- [72] Jafry Y., Schuermann T. (2004). "Measurement, Estimation and Comparison of Credit Migration Matrices", Journal of Banking & Finance, doi:10.1016/j.jbankfin.2004.06.004
- [73] Jafry Y., Schuermann T. (2003). "Measurement and Estimation of Credit Migration Matrices", The Wharton Financial Institutions Center
- [74] Johansen S. (1973). "A Central Limit Theorem for Finite Semigroups and its Application to the Imbedding Problem for Finite State Markov Chains", Z. Wahrscheinlichkeitstheorie verw. Geb. 26, pp 171-190
- [75] Culver W. J. (1966). "On the Existence and Uniqueness of the Real Logarithm of a Matrix", Proceedings of the American Mathematical Society, Vol. 17, No. 5, pp 1146-1151
- [76] Cardoso J. R. (2004). "An Explicit Formula for the Matrix Logarithm", General Mathematics, 15A18, eprint arXiv:math/0410556
- [77] Helton B. W. (1968). "Logarithms of Matrices", Proceedings of the American Mathematical Society, Vol. 19, No. 3, pp 733-738
- [78] Geweke J., Marshall R., Zarkin G. (1986). "Mobility Indices in Continuous Time Markov Chains", Econometrica, Vol. 54, No. 6, pp 1407-1423