



UNIVERSIDAD NACIONAL AUTÓNOMA DE  
MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN



PROCESOS ESTOCÁSTICOS

Y

REDES DE PETRI ESTOCÁSTICAS

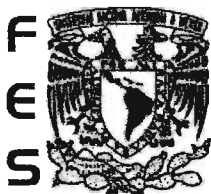
T E S I S

QUE PARA OBTENER EL TÍTULO DE:

A C T U A R I O

P R E S E N T A:

MAYRA LORENA DÍAZ SOSA



ACATLÁN

ASESOR: DR. ZBIGNIEW OZIEWICZ

ACATLÁN, EDO. DE MÉXICO. SEPTIEMBRE DE 2005.

m1348745



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**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.

Tesis de Licenciatura financiada por la Dirección General de Asuntos del Personal Académico de la Universidad Nacional Autónoma de México a través del Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), clave IN-105402; y por el Consejo Nacional de Ciencia y Tecnología (CONACyT) de México, clave U 41214 F.

**ESTA TESIS NO SALE  
DE LA BIBLIOTECA**

# A mis padres.

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recopial.

NOMBRE: Mayra Lorena Díaz Sosa

FECHA: 27 octubre - 2005

FIRMA: 

Con todo mi cariño, agradecimiento y admiración.

Mayra Lorena Díaz Sosa.



*Y a todo aquel que consulte  
este trabajo con la  
esperanza de que le sirva de algo.*



# Agradecimientos

A Dios, por dármele todo aun sin merecerlo.

A mis padres: María Concepción Sosa Tuñón y Raúl Díaz Nájera, a quienes debo lo que soy, lo que fui y lo que seré. Gracias por su amor y por señalarme el buen camino. Ustedes han sido una lección de vida que nunca olvidaré. Los amo mucho.

Gracias a mi novio Víctor Manuel Rangel Cortés por darme todo su amor y su apoyo incondicional, por comprender la locura de mi cabeza y de mi corazón. Te amo con todo mi ser.

A la Universidad Nacional Autónoma de México, a la Facultad de Estudios Superiores Acatlán y a todos mis profesores, por iluminar mi vida con conocimientos. En especial, gracias a mi asesor, el Dr. Zbigniew Oziewicz, por todo su apoyo y su infinita paciencia. Al Lic. Víctor Manuel Ulloa Arellano por ser uno de mis mejores maestros y además mi amigo. Al Dr. Félix Mendoza Martínez por toda su ayuda, su calidad humana y por recordarme quién soy... ¡Dios bendiga a la lengua rusa!

A la Dra. Mónica A. Limones Valencia por ayudarme a superar mis miedos y a conocerme un poquito mejor.

A todos mis amigos, por apoyarme en todo momento y por siempre creer en mí. Su cariño me impulsó a continuar semestre tras semestre.

A todos y cada uno de ustedes, Dios los bendiga.



## VIII

Gracias al CONACyT y al PAPIIT por el valioso soporte para la elaboración de este proyecto.

# Resumen

En la presente Tesis primeramente se establecen algunas de las más importantes definiciones en materia de probabilidad y de teoría estocástica, enfocándose particularmente al caso de los procesos estocásticos conocidos como cadenas y procesos markovianos. Posteriormente se presentan las principales características de las redes de Petri y se señala a las mismas como una herramienta matemática útil para la construcción, análisis formal, y representación gráfica de sistemas. Por último, se describe la forma en que las redes de Petri y los procesos markovianos pueden aplicarse de manera conjunta en el estudio de sistemas aleatorios.



# Objetivos

## Objetivo general

Iniciar una nueva línea de investigación dirigida a la aplicación de redes de Petri en la ciencia actuarial al presentar, explicar y vincular los conceptos fundamentales de la teoría estocástica y de la teoría de redes de Petri.

## Objetivos particulares

- Construir la definición de probabilidad partiendo del concepto de álgebra booleana y álgebra booleana de conjuntos.
- Difundir las redes de Petri como herramientas formales que permiten el análisis cualitativo y cuantitativo de sistemas.
- Proporcionar métodos alternativos para el estudio de sistemas que involucren aleatoriedad.



# Delimitación

Dentro de la teoría de los procesos estocásticos, el trabajo se enfocará al caso de aquellos que se ajusten a la teoría markoviana.



# Hipótesis

- Existen beneficios para la carrera de Actuaría si las teorías markoviana y de redes de Petri se estudian de manera conjunta.
- La teoría markoviana puede enriquecer a la teoría de las redes de Petri, y viceversa.





# Justificación

Los procesos estocásticos, específicamente los procesos markovianos, tienen una importante aplicación en las ciencias actuariales, matemáticas e ingeniería, principalmente. Su estudio es extenso y requiere de una amplia gama de teorías matemáticas.

Actualmente, las redes de Petri han empezado a jugar un papel preponderante, como herramientas de modelado matemático al ser formalismos que permiten la validación cualitativa y cuantitativa de sistemas.

En virtud de lo anterior y tomando en cuenta que la aplicación de las redes de Petri al campo de los procesos markovianos es incipiente, es de especial importancia iniciar una línea de investigación dirigida a proporcionar métodos alternativos para el estudio de sistemas que involucren aleatoriedad, ya que ésta se encuentra presente en las diversas áreas en las que el actuario ejerce su profesión.



# Introducción

Una de las principales problemáticas de nuestro tiempo es la creciente complejidad de la estructura y funcionamiento de sistemas modernos, por lo que se requiere de nuevos métodos formales apropiados para la especificación de sistemas y para su validación tanto cualitativa como cuantitativa. Las redes de Petri son formalismos útiles en el análisis de sistemas modernos ya que permiten describirlos de manera concisa y apropiada debido a que ofrecen una representación gráfica ordenada, versátil y lógica de los módulos que componen a un sistema, así como de la interacción que guardan éstos entre sí. Las redes de Petri fueron creadas en 1962 por Carl Adam Petri para su Tesis doctoral, en la cual describe sus fundamentos teóricos. Desde entonces, cada vez son más los estudiosos de este tema y han surgido muchas variedades de las mismas en los últimos años.

Por otra parte, cuando la aleatoriedad se manifiesta en un sistema la complejidad del mismo aumenta considerablemente, por lo que es necesario el uso de herramientas probabilísticas que permitan su correcta modelación y estudio. La teoría de la probabilidad se inició prácticamente con el análisis matemático de los juegos de azar realizado primero por los matemáticos Pierre du Fermat (1601-1665) y Blas Pascal (1623-1662). Christian Huygens (1629-1695) publicó en 1657 el primer tratado sobre problemas relacionados con juegos de azar, el cual sirvió como base para el gran desarrollo experimentado por esta teoría durante el siglo XVIII, con el aporte de Jacobo Bernoulli (1654-1705) y Pierre Simon Laplace (1749-1827). Fueron Markov (1856-1922) y Kolmogorov (1903-1987) dos de los principales matemáticos que contribuyeron al estudio de la aleatoriedad a través de modelos conocidos como procesos estocásticos, los cuales permiten modelar una amplia gama de fenómenos, bajo un sólido fundamento matemático.

Sin embargo, las redes de Petri y los procesos estocásticos no han sido difundidos en la debida proporción debido a su reciente origen, en el primer caso; y a la complejidad de los cálculos que implican, en el segundo. Y menos aún han sido consideradas las ventajas de estudiarlos y aplicarlos de manera conjunta, prescindiendo así del uso de herramientas matemáticas potencialmente útiles.

El hecho de que cada vez son más los actuarios que se desarrollan profesionalmente no sólo en las áreas tradicionales de seguros, finanzas, estadística y economía, sino también en las áreas de sistemas computacionales, industriales, organizacionales y logísticos entre otros, me motivó a escribir la presente Tesis con la intención de iniciar una línea de investigación dirigida a la aplicación de redes de Petri en la ciencia actuarial, ya que en el anterior plan de estudios de la carrera de Actuaría y en el de reciente implantación no existe ninguna asignatura dirigida al estudio de las redes de Petri en combinación con los procesos estocásticos.

En este sentido, la presente Tesis es interdisciplinaria y ofrece un enfoque novedoso acerca de estas dos herramientas para el estudio de sistemas estocásticos, por lo que me permito puntualizar algunos de los principales logros de este trabajo:

- Para entender el concepto de proceso estocástico es indispensable definir primero el concepto de probabilidad. Sin embargo, el concepto de probabilidad no puede ser entendido sin el álgebra booleana propuesta por el matemático inglés George Boole en 1854. Al revisar varios libros de procesos estocásticos noté que la mayoría de ellos ni siquiera mencionan el álgebra de Boole. Algunos autores, como Durrett 1999 [13], Papoulis 1965 [26], Parzen 1962 [27] y Ross 1996 [30], hacen un breve repaso de probabilidad a manera de introducción a los procesos estocásticos pero sin considerar el álgebra booleana. Algunos otros autores, como Vélez 1996 [37], ni siquiera hacen un repaso de probabilidad. Debido a lo anterior y a la falta de material que provea un orden lógico que nos permita entender el concepto de probabilidad, uno de los logros de esta Tesis es establecer la definición de probabilidad bajo el enfoque del álgebra de Boole.
- En la mayoría de los libros de probabilidad los conceptos de probabilidad y de variable aleatoria se estudian por separado y no se explica la relación

que existe entre ellos. Este trabajo explica que el álgebra de variables aleatorias satisface las leyes del álgebra de Boole y que tanto el concepto de álgebra de variables aleatorias como el de probabilidad se desprenden del álgebra booleana.

- Por otra parte, se explica que el nombre de variable aleatoria puede ser engañoso, pues la definición clásica de este concepto matemático establece que en realidad se trata de una función numérica y no de una variable. Sin embargo, existe un enfoque alternativo, y poco conocido, que propone que la variable aleatoria es, en efecto, variable. Este trabajo introduce este enfoque alternativo, por lo que cumple una labor de difusión.
- En algunas ocasiones la notación utilizada en libros sobre procesos estocásticos es muy complicada y puede causar confusión. En esta Tesis se establecen algunas definiciones fundamentales de las teorías estocástica y probabilista y de las redes de Petri, empleando una notación accesible e incluyendo más de 20 ejemplos, 18 de ellos inéditos, que ilustran el significado de los conceptos y que guardan un orden lógico como propuesta de enseñanza, lo cual resulta una aportación pedagógica de este trabajo.
- Cabe destacar que el tema de estudio de la presente Tesis lo he presentado y expuesto con éxito en *The First International Conference on Nonlinear, Fractional and Stochastic Dynamics*, celebrado en la Facultad de Estudios Superiores Cuautitlán del 24 al 28 de enero de 2005, con el nombre de *Stochastic Processes and Operads of Graphs*. La memoria sobre mi participación en dicho Congreso titulada *Some Basic Notions of Probability and Markov Chains*, la cual actualmente se encuentra en proceso de publicación, se incluye en el Apéndice de la presente Tesis.
- La gran cantidad de simbología matemática necesaria en la mayoría de los textos científicos dificulta su composición en editores de texto convencionales. A pesar de su complejidad, el lenguaje  $\text{\LaTeX} 2_{\epsilon}$  ofrece una gran calidad tipográfica para textos científicos por lo que las más importantes casas editoriales, como Springer-Verlag, lo emplean y los principales congresos internacionales y revistas científicas arbitradas lo exigen en la presentación de proyectos. Cabe destacar que la Actuaría y Maestra en Educación, MariCarmen González Videgaray, ha sido una de las principales promotoras del uso de esta importante herramienta en nuestro Campus. En este proyecto se empleó el lenguaje  $\text{\LaTeX} 2_{\epsilon}$ , razón por la

Campus. En este proyecto se empleó el lenguaje  $\text{\LaTeX} 2_{\epsilon}$ , razón por la cual una de las conclusiones de este proyecto es el impulso a la utilización de dicho lenguaje por los estudiantes de las áreas de matemáticas para la elaboración de sus trabajos de investigación, así como para el desarrollo de sus Tesis de Licenciatura y de Posgrado.

En el Capítulo 1, *Teoría estocástica*, se efectúa una revisión sobre conceptos fundamentales de la teoría de la probabilidad. Específicamente, en este capítulo se enfatiza la importancia del álgebra booleana como concepto matemático indispensable para comprender la teoría de la probabilidad y se estudia la relación entre los conceptos de variable aleatoria y probabilidad. Además, se habla de los conceptos de aleatoriedad y proceso estocástico. Este capítulo brinda los elementos fundamentales para la cabal comprensión de la teoría markoviana.

En el Capítulo 2, *Teoría markoviana*, se hace un estudio extenso de los procesos estocásticos conocidos como cadenas de Markov así como de los procesos de Markov, ya que ambos permiten modelar una amplia gama de fenómenos de interés para el actuario, como son la ocurrencia de siniestros, índices de contaminación, índices inflacionarios, fluctuación cambiaria y fenómenos demográficos, entre otros. En este capítulo se crearon algunos ejemplos con el fin de ilustrar las definiciones y utilidad de los teoremas del mismo.

El Capítulo 3, *Redes de Petri*, se enfoca esencialmente a la descripción de las principales características de las redes de Petri ordinarias y señala a las mismas como una herramienta matemática útil en la construcción, análisis formal, y representación gráfica de sistemas. Al igual que en el capítulo anterior, se crearon algunos ejemplos.

En el Capítulo 4, *Redes de Petri estocásticas*, se describe la forma en que las redes de Petri y la teoría markoviana pueden aplicarse de manera conjunta en el estudio de sistemas aleatorios, idea expuesta por primera vez de forma independiente por S. Natkin y M. K. Molloy a inicios de los años ochenta. Se proponen algunas aplicaciones en áreas en donde el actuario tiene oportunidades de desarrollo potenciales.

Finalmente, se presentan las conclusiones y recomendaciones.

Esta Tesis fue posible gracias a la atinada dirección del Dr. Zbigniew Ozie-wicz, eminente catedrático e investigador de la máxima casa de estudios, quien ha realizado importantes aportaciones matemáticas, especialmente en el área de los lenguajes gráficos, como son los Operades y sus aplicaciones. Estas investigaciones proporcionan una importante alternativa al lenguaje científico convencional.





# Índice general

<b>Introducción</b>	<b>1</b>
<b>1. Teoría estocástica</b>	<b>7</b>
1.1. Álgebra de Boole . . . . .	9
1.2. Probabilidad . . . . .	13
1.2.1. Teorema de Bayes . . . . .	15
1.3. Espacio muestral . . . . .	18
1.4. Álgebra de variables aleatorias . . . . .	21
1.4.1. Relación entre probabilidad y álgebra de variables aleatorias . . . . .	24
1.4.2. Variables aleatorias discretas . . . . .	27
1.4.3. Variables aleatorias continuas . . . . .	28
1.4.4. Momentos de una variable aleatoria . . . . .	29
1.4.5. Distribución conjunta de variables aleatorias . . . . .	30
1.5. Proceso estocástico . . . . .	31

<b>2. Teoría markoviana</b>	<b>35</b>
2.1. Cadenas de Markov . . . . .	35
2.1.1. Clasificación de los estados . . . . .	43
2.1.2. Distribución de estado estable . . . . .	50
2.1.3. Cadenas absorbentes y comportamiento transitorio . . . . .	55
2.2. Procesos de Markov . . . . .	60
2.2.1. Distribución de estado estable . . . . .	66
<b>3. Redes de Petri</b>	<b>69</b>
3.1. Representación gráfica . . . . .	70
3.2. Estructura . . . . .	71
3.3. Comportamiento Dinámico . . . . .	76
3.4. Propiedades . . . . .	80
3.5. Análisis de las redes de Petri . . . . .	82
3.5.1. Análisis del conjunto alcanzable . . . . .	82
3.5.2. Análisis invariante . . . . .	83
<b>4. Redes de Petri estocásticas</b>	<b>95</b>
4.1. Posibles aplicaciones . . . . .	103
4.1.1. Modelos de crecimiento poblacional . . . . .	103
4.1.2. Comunicación y control . . . . .	104

4.1.3. Administración e Investigación de Operaciones . . . . .	104
4.1.4. Simulación . . . . .	106
<b>Conclusiones y recomendaciones</b>	<b>109</b>
<b>Apéndice</b>	<b>111</b>
<b>Bibliografía</b>	<b>125</b>
<b>Índice alfabético</b>	<b>129</b>



# Índice de figuras

1.1. Probabilidad y álgebra de variables aleatorias . . . . .	27
2.1. Cadena de Markov ergódica . . . . .	39
2.2. Cadena reducible . . . . .	45
2.3. Cadena de Ehrenfest con $N = 3$ . . . . .	52
3.1. Red de Petri . . . . .	73
3.2. Secuencia de descargas $\sigma = t_1 t_1 t_3 t_3$ . . . . .	79
3.3. Red de Petri no acotada . . . . .	79
3.4. Árbol del conjunto alcanzable . . . . .	84
3.5. Grafo del conjunto alcanzable . . . . .	85
4.1. Redes de Petri estocásticas y teoría markoviana . . . . .	96
4.2. Red de Petri estocástica . . . . .	99
4.3. Grafo del conjunto alcanzable de la red de Petri estocástica . . . . .	101
4.4. Proceso de Markov de la red de Petri estocástica . . . . .	102



# Capítulo 1

## Teoría estocástica

*“Es a esta ignorancia a la que se le da el nombre de azar; pero este bautismo no sería más que una pirueta si este azar no pudiera calcularse y si la teoría no permitiera ninguna aplicación práctica”.*

**Jean-Louis Boursin**

Para establecer qué es la teoría estocástica, y comprender su importancia, es preciso definir primero una serie de conceptos.

Primero nos ocuparemos de la definición de proceso. *Proceso* es el sistema que representa una serie de acciones, o bien las fases de un fenómeno natural o de una operación artificial, que tienen lugar continua, sucesiva o regularmente<sup>1</sup>.

Proceso estocástico es aquel cuya evolución en el tiempo es aleatoria.

Pero, ¿qué es la aleatoriedad? Y si existe la aleatoriedad, debe existir la no aleatoriedad.

A la no aleatoriedad se le conoce como *determinismo* y es la teoría que supone que la evolución de los fenómenos naturales o de otra índole está completamente determinada por las condiciones iniciales del mismo.

---

<sup>1</sup>*Diccionario de Términos Científicos y Técnicos*. Barcelona, McGraw Hill, Vol. IV, 1981, p. 1629.



A pesar de que en el Diccionario de la Lengua Española las definiciones de azar y aleatoriedad redundan en la casualidad y lo fortuito, Steen 1998 [32] sostiene que se llama *aleatoriedad* a la cualidad que poseen algunos fenómenos de tener resultados individuales *inciertos* pero con un patrón regular de los mismos en muchas repeticiones. Es decir, que aleatorio no es propiamente sinónimo de fortuito, sino la descripción de un tipo de orden diferente del determinista que suele asociarse a la ciencia y las matemáticas.

Mucho se ha discutido sobre la existencia del azar. Mario Bunge afirma que el azar existe en la mecánica cuántica mientras que otros, como Henri Poincaré, sostienen que a lo que llamamos azar no es otra cosa que a la ignorancia derivada de la falta de información sobre un sistema. En fin, este tema ha provocado polémica y ha sido causa de múltiples debates a lo largo de los últimos años.

Afirmar que el universo es aleatorio o determinista resulta sumamente aventurado. Menos ambicioso, pero más útil, resulta modelar un subsistema del mundo real por medio de un sistema matemático aleatorio o determinista.

Nosotros nos limitaremos a hacer la siguiente distinción. Consideraremos a un subsistema del universo como determinista si, ignorando efectos externos inesperados, cada vez que lo devolvemos a lo que aparentemente es el mismo estado inicial hace lo mismo durante un período de tiempo no nulo. En cambio, lo consideraremos aleatorio si estados inicialmente indistinguibles pueden conducir inmediatamente a resultados muy diferentes. Una vez establecido lo anterior, podremos decidir si adoptar un modelo determinista o aleatorio para el estudio del subsistema del universo.

Al conjunto de ecuaciones cuyo propósito es describir el comportamiento de un sistema se le llama *modelo matemático*. Al comparar observaciones del mundo real con algún modelo concreto, sólo del modelo podremos decir con seguridad que es aleatorio o determinista. Y si es lo uno o lo otro, entonces aquellos aspectos que éste recoge del mundo real también lo serán.

La teoría matemática que sustenta a los modelos aleatorios es conocida como *teoría estocástica*, la cual ha cobrado cada vez mayor fuerza debido a que permite construir modelos muy cercanos a la realidad.

Sin embargo, los modelos aleatorios no han sido explotados en la debida pro-

porción, en parte debido a la complejidad de los cálculos que éstos implican.

Para efectos de este texto, a los modelos aleatorios los llamaremos procesos estocásticos.

¿Por qué los procesos estocásticos resultan de especial interés para el actuario? Recordemos que actuario es el especialista capacitado en la evaluación, administración y prevención de riesgos en áreas de seguros, finanzas, estadística, demografía, sistemas, entre otras.

Por ejemplo, el comportamiento del mercado y de los principales indicadores económicos, las primas que debe cobrar una compañía aseguradora a sus clientes así como la reserva con la que debe contar para cubrir sus compromisos, la evaluación de proyectos, la optimización ya sea de recursos o de tiempo, son eventos que involucran riesgos en cuyo análisis suele interesarse el actuario debido a su campo de acción.

El riesgo implica incertidumbre y una mayor o menor comprensión de los procesos estocásticos le permitirá al actuario en la misma medida tomar en cuenta el papel del azar para cumplir con su misión.

## 1.1. Álgebra de Boole

George Boole (1815-1864), lógico y matemático inglés nacido en Lincoln y fallecido en Cork (Irlanda), comenzó sus estudios sobre matemáticas por su cuenta en el año de 1835. El propósito de Boole era reducir las “Leyes del Pensamiento” a una rama axiomática del álgebra. El título de su famoso libro en la materia es *An Investigation of the Laws of Thought, on which are founded the Mathematical Theories of Logic and Probability*, publicado en 1854. En la introducción de dicho trabajo, Boole enfatizó las diferencias entre la lógica y el álgebra, y explicó que intentaba unificar a estas dos ramas de las matemáticas.

Para producir un álgebra de proposiciones, Boole se basó en las dos operaciones binarias  $\vee$ ,  $\wedge$  y la operación unaria  $'$ , las cuales satisfacen leyes (tales como la asociativa, conmutativa y distributiva) que se cumplen en la mayoría de las estructuras algebraicas, como es el caso de los grupos y anillos.

**1.1 Definición.** *Retículo* es la terna  $(R, \vee, \wedge)$  donde  $R$  es un conjunto y  $\vee, \wedge$  son dos operaciones binarias definidas en  $R$  que cumplen las siguientes propiedades:

- *Leyes asociativas:*  $x \vee (y \vee z) = (x \vee y) \vee z$  y  $x \wedge (y \wedge z) = (x \wedge y) \wedge z$ .
- *Leyes conmutativas:*  $x \vee y = y \vee x$  y  $x \wedge y = y \wedge x$ .
- *Leyes idempotentes:*  $x \vee x = x$  y  $x \wedge x = x$ .
- *Leyes de absorción:*  $x \vee (x \wedge y) = x = x \wedge (x \vee y)$ .

Los retículos poseen las siguientes propiedades:

- El retículo es *acotado* si posee máximo y mínimo. Se designa por  $1$  al máximo y por  $0$  al mínimo.
- Sea  $R$  un retículo acotado. Dado  $x \in R$  se dice que  $x' \in R$  es *complementario* de  $x$  si  $x \vee x' = 1$  y  $x \wedge x' = 0$ . Un retículo es *complementario* si todos sus elementos poseen complementario.
- Un retículo es *distributivo* si para cualesquiera  $x, y, z \in R$  se cumple que:  $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$  y  $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ .

**1.2 Definición.** *Álgebra de Boole* es el retículo  $\mathcal{B}$  con la operación unaria  $'$  y dos constantes,  $0$  y  $1$ , que satisfacen las siguientes axiomas:

- *Leyes de identidad:*  $x \vee 0 = x$  y  $x \wedge 1 = x$ .
- *Leyes de Morgan:*  $(x \vee y)' = x' \wedge y'$  y  $(x \wedge y)' = x' \vee y'$ .

En otras palabras, álgebra de Boole es un retículo complementario y distributivo.

A los elementos de un álgebra booleana se les denomina simplemente como *eventos*.

**1.3 Definición.** Si  $A$  es un conjunto finito, el número de elementos en  $A$  es un entero  $N$ , llamado *cardinalidad*.

**1.4 Definición.** Para cualquier conjunto  $A$ , el conjunto de todos los subconjuntos de  $A$  recibe el nombre de *conjunto potencia* de  $A$ . Si  $A$  es un conjunto finito con cardinalidad  $N$ , entonces la cardinalidad del conjunto potencia de  $A$  es  $2^N$ . Por esta razón, el conjunto potencia se escribe como  $2^A$ .

**1.5 Definición.** *Álgebra booleana de conjuntos* define todas las operaciones  $W = 2^U$  del conjunto universal  $U$  de acuerdo con las siguientes reglas:

- $x \vee y = x \cup y$ .
- $x \wedge y = x \cap y$ .
- $x' = U \setminus x = x^c$ .
- $1 = U, 0 = \emptyset$ .

**1.6 Definición.** Sea  $Q \subset 2^U$ , es decir, una familia de conjuntos. *Función de conjunto* es aquella cuyo dominio es una familia de conjuntos y cuyo recorrido es un subconjunto de los números reales

$$f : Q \rightarrow \mathbb{R}$$

**1.7 Definición.** Una función de conjunto es de *aditividad finita* si para cada par de elementos  $A$  y  $B$ , de la familia, y cuya unión es elemento de  $Q$ , se cumple que

$$f(A \cup B) = f(A) + f(B) - f(A \cap B)$$

De las definiciones anteriores se obtienen los siguientes resultados.

Por las leyes de Morgan tenemos que  $(G \cap H)^c = G^c \cup H^c$ , entonces:

$$(A^c \cup B^c)^c = A \cap B$$

y como se tiene que si  $A$  y  $B$  pertenecen a  $Q$ , y el complemento de cada conjunto de  $Q$  igualmente pertenece a  $Q$ , entonces:

$$A^c \cup B^c \in Q \text{ y } (A^c \cup B^c)^c \in Q$$

luego

$$A \cap B \in Q$$

**1.8 Definición.**  $A - B = A \cap B^c$ .

Si  $A$  y  $B$  pertenecen a  $Q$ , entonces como  $A - B = A \cap B^c$  y dado que  $B^c$  también pertenece a  $Q$ , entonces se tiene que:

$$A - B \in Q$$

Por tanto, si  $Q$  es cerrada para la unión, intersección y diferencia (complementación) de conjuntos entonces  $Q$  es álgebra de Boole de conjuntos.

Entonces, como  $A \cup A^c = U$ , se tiene que  $U \in Q$ , y también como  $A - A = \emptyset$ , para cada  $A \in Q$ ,

$$\emptyset \in Q$$

Por consiguiente,  $Q_0 = \{\emptyset, U\}$  es la menor álgebra de Boole que se puede construir a partir de los subconjuntos de un conjunto universal  $U$ .  $Q_k$  es el conjunto de *todos* los subconjuntos de  $U$  y es, por tanto, el mayor álgebra de Boole que se puede construir con subconjuntos de  $U$ .

Si  $f$  es una función de conjunto con aditividad finita definida sobre un álgebra de Boole  $Q$ , entonces si  $A$  y  $B$  son elementos cualesquiera de  $Q$ , se tiene:

$$f(A \cup B) = f(A \cup (B - A)) = f(A) + f(B - A) \quad (1.1)$$

por ser  $f$  con aditividad finita y los conjuntos  $A$  y  $B - A$  disjuntos. Luego

$$f(A \cup B) = f(A) + f(B - A) \quad (1.2)$$

Y también:

$$\begin{aligned} A &= A \cap U \\ &= A \cap (B \cup B^c) \\ &= (A \cap B) \cup (A \cap B^c) \end{aligned}$$

entonces:

$$f(A) = f(A \cap B) + f(A \cap B^c)$$

y

$$f(A \cap B^c) = f(A) - f(A \cap B)$$

Y, además,  $A \cup B = (A \cap B^c) \cup B$ , con  $(A \cap B^c) \cap B = \emptyset$ ; por consiguiente:

$$f(A \cup B) = f(A \cap B^c) + f(B) \quad (1.3)$$

Sustituyendo ( 1.2) en ( 1.3), se obtiene la definición 1.7:

$$f(A \cup B) = f(A) + f(B) - f(A \cap B) \quad (1.4)$$

**1.9 Definición.** Una función de conjunto con aditividad finita, definida no negativa se llama *medida con aditividad finita* o, simplemente, una *medida*.

Una medida  $f$  sobre un álgebra de Boole  $Q$  cumple las propiedades siguientes:

1. Para cada par de elementos  $A$  y  $B$  de  $Q$ :

$$f(A \cup B) \leq f(A) + f(B) \quad (1.5)$$

lo cual se concluye directamente de ( 1.4).

2. Si  $A$  es subconjunto de  $B$  entonces  $f(B - A) = f(B) - f(A)$ , ya que por ( 1.1),  $f(B - A) = f(A \cup B) - f(A)$ , pero si  $A$  es subconjunto de  $B$ , entonces  $A \cup B = B$ .
3. Si  $A$  es subconjunto de  $B$  y dado que  $f$  es medida, entonces se tiene:  $f(B - A) = f(B) - f(A)$ ; por tanto:

$$f(A) \leq f(B)$$

4. Como  $A - A = \emptyset$ , entonces:

$$f(\emptyset) = f(A - A) = f(A) - f(A) = 0$$

## 1.2. Probabilidad

La teoría de la probabilidad se inició prácticamente con el análisis matemático de los juegos de azar realizado primero por los matemáticos Pierre du Fermat (1601-1665) y Blas Pascal (1623-1662). Christian Huygens (1629-1695)

publicó en 1657 el primer tratado sobre problemas relacionados con juegos de azar, el cual sirvió como base para el gran desarrollo experimentado por esta teoría durante el siglo XVIII, con el aporte de Jacobo Bernoulli (1654-1705) y Pierre Simon Laplace (1749-1827), y posteriormente con las contribuciones de Chebyshev, Markov y Kolmogorov; este último estableció en 1933 la presentación axiomática que construye la base de la moderna teoría de la probabilidad.

Con todos los elementos previamente señalados ahora podemos construir en forma axiomática el concepto de probabilidad.

**1.10 Definición.** Sea  $Q$  un álgebra de Boole. *Probabilidad* de un evento  $A_i \in Q$ , es la función  $P$  que satisface los axiomas:

1.  $0 \leq P(A_i) \leq 1$ .
2.  $P(\Omega) = 1$  donde  $\Omega = \bigcup_{A_i \in Q} A_i$ .
3.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  donde  $B \in Q$ .
4.  $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$ , donde  $A_i \cap A_j = \emptyset$  e  $i \neq j$ .

De manera que espacio de probabilidad es toda dupla  $(Q, P)$ .

Equivalentemente, una función  $P$  definida en  $Q$  se llama *probabilidad* si:

1.  $P$  es no negativa.
2.  $P(\Omega) = 1$ , esto es, una medida de probabilidad es una medida que asocia 1 a la unión de todos los eventos en  $Q$ .
3.  $P$  es de aditividad finita.
4.  $P$  es de *aditividad numerable*. Al escribir  $n$  se entiende que  $Q$  es un conjunto de eventos finito. En el caso de que  $Q$  fuera infinito escribiríamos  $\infty$  en lugar de  $n$ . Más adelante ahondaremos sobre la aditividad numerable.

De la definición de probabilidad se desprende lo siguiente:

**1.1 Teorema.** Para cada evento  $A$ ,  $P(A^c) = 1 - P(A)$ .

*Demostración.* Como

$$\Omega = A \cup (A^c) \text{ y } A \cap (A^c) = \emptyset$$

entonces

$$P(\Omega) = P(A) + P(A^c) = 1$$

y

$$P(A^c) = 1 - P(A)$$

□

**1.2 Teorema.**  $P(\emptyset) = 0$ .

*Demostración.* Sea  $\{A_n\} = \{\Omega, \emptyset, \emptyset, \dots\}$ . Entonces por las condiciones 2 y 4 de la definición 1.10:

$$P\left(\bigcup A_n\right) = P(\Omega) = P(\Omega) + P(\emptyset) + P(\emptyset) + \dots$$

Es decir

$$P(\emptyset) + P(\emptyset) + \dots = 0$$

por lo que  $P(\emptyset) = 0$ .

□

Si  $A$  y  $B$  son dos eventos, entonces

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Este resultado, al igual que los anteriores, se desprende de las propiedades demostradas antes para una función de conjunto con aditividad finita, y dado que la probabilidad es una función de conjunto con aditividad finita definida no negativa, entonces los resultados obtenidos para estas funciones son aplicables a la probabilidad.

### 1.2.1. Teorema de Bayes

Para comprender la teoría estocástica, se requiere tener conocimiento sobre algunos conceptos fundamentales de teoría de la probabilidad ya que ésta es



la rama de las matemáticas que describe la aleatoriedad. Esta subsección se ocupa de una introducción básica sobre el tema. Para un estudio exhaustivo sobre la materia se recomiendan al lector los libros de probabilidad de Devore 1998 [11], Lipschutz 1991 [22], Mendenhall 1994 [23] y Papoulis 1965 [26], los tres primeros con un enfoque práctico y el último con un nivel teórico más elevado.

**1.11 Definición.** Si dos eventos  $A$  y  $B$  son tales que

$$A \cap B = \emptyset$$

entonces los dos eventos son *mutuamente excluyentes* o *disjuntos*.

**1.12 Definición.** *Conjunto de eventos exhaustivos*, es el conjunto de eventos cuya unión forma el conjunto  $\Omega$  de todos los eventos de un álgebra de Boole.

**1.13 Definición.** Reciben el nombre de *eventos mutuamente excluyentes exhaustivos* los eventos  $\{A_1, A_2, \dots, A_n\}$  tales que

$$A_i A_j = A_i \cap A_j = \emptyset \quad \text{para toda } i \neq j$$

$$\Omega = A_1 \cup A_2 \cup \dots \cup A_n \tag{1.6}$$

**1.14 Definición.** La *probabilidad condicional* de un evento  $A$ , dado que el evento  $B$  ocurrió se denota como  $P[A|B]$  y se define como

$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$

siempre que  $P[B] \neq 0$ .

La *independencia estadística* de eventos puede definirse como sigue. Dos eventos  $A$  y  $B$  se dicen estadísticamente independientes si

$$P[A \cap B] = P[A]P[B] \tag{1.7}$$

Para tres eventos  $A$ ,  $B$  y  $C$  estadísticamente independientes, cada par de eventos debe satisfacer (1.7). De manera que

$$P[A \cap B \cap C] = P[A]P[B]P[C]$$

y así sucesivamente.

De donde se desprende que, para dos eventos independientes  $A$  y  $B$

$$P[A|B] = \frac{P[A \cap B]}{P[B]} = P[A]$$

lo cual significa que la ocurrencia de un evento  $B$  no afecta la probabilidad de ocurrencia del evento independiente  $A$  en forma alguna y viceversa.

El siguiente teorema es comúnmente conocido como el *teorema de probabilidad total*.

**1.3 Teorema** (de probabilidad total). *Considere un evento  $B$  y el conjunto de eventos mutuamente excluyentes exhaustivos  $\{A_1, A_2, \dots, A_n\}$ . Si el evento  $B$  ocurre, debe ocurrir conjuntamente con exactamente uno de los eventos mutuamente excluyentes exhaustivos  $A_i$ . Esto es*

$$P[B] = \sum_{i=1}^n P[A_i \cap B]$$

De la definición de probabilidad condicional podemos escribir

$$\begin{aligned} P[A_i \cap B] &= P[A_i|B]P[B] \\ &= P[B|A_i]P[A_i] \end{aligned}$$

lo que nos permite obtener una segunda forma del teorema de probabilidad total

$$P[B] = \sum_{i=1}^n P[B|A_i]P[A_i]$$

Esta última ecuación sugiere que para encontrar la probabilidad de algún evento complejo  $B$ , es posible simplificar el cálculo de la misma condicionando el evento  $B$  a algún evento  $A_i$  de forma tal que, al calcular la probabilidad del evento  $B$  dado el evento  $A_i$  y después multiplicarla por la probabilidad del evento condicional  $A_i$ , se obtenga la probabilidad  $P[A_i B]$ . Una vez hecho esto para un conjunto de eventos mutuamente excluyentes exhaustivos  $A_i$  procederemos a sumar estas probabilidades para encontrar la probabilidad de ocurrencia del evento  $B$ .

Ahora tenemos todos los elementos necesarios para comprender el *teorema de Bayes*.

El teólogo y matemático inglés Thomas Bayes (1702-1761) hizo importantes contribuciones en materia de probabilidad y estadística y sus ideas han creado controversia entre muchos matemáticos durante años.

Probablemente, el escrito más famoso de Bayes fue *Essay Towards Solving a Problem in the Doctrine of Chances* publicado por la Royal Society of London en 1764. En dicho ensayo se describe la técnica estadística de Bayes hoy conocida como *estimación bayesiana*, técnica que a nuestros días sigue siendo tema de discusión entre matemáticos de todo el mundo. El famoso teorema que hoy lleva su nombre fue publicado en 1763 y es considerado una de sus principales contribuciones matemáticas.

**1.4 Teorema (Bayes, 1763).** Sea  $\{A_i\}$  un conjunto de eventos mutuamente excluyentes y exhaustivos. Entonces

$$P[A_i|B] = \frac{P[B|A_i]P[A_i]}{\sum_{j=1}^n P[B|A_j]P[A_j]} \quad (1.8)$$

Este teorema nos permite calcular la probabilidad de un evento condicionado a un segundo al calcular la probabilidad del segundo condicionado a el primero y otros términos.

### 1.3. Espacio muestral

La definición de probabilidad ha sido construida hasta el momento únicamente a través del álgebra de Boole. En esta sección se extiende esta definición incluyendo el concepto de espacio muestral y se explica brevemente la relación entre éste y el álgebra de Boole.

**1.15 Definición.** *Espacio muestral* asociado a un experimento aleatorio, es el conjunto de todos los resultados posibles de dicho experimento.

Cada uno de los elementos de un espacio muestral  $\Omega$  se llama *suceso elemental*. Un *suceso* es un subconjunto del espacio muestral, es decir, un subconjunto de sucesos elementales del experimento aleatorio.

**1.16 Definición.** *Espacio muestral finito* asociado a un experimento aleatorio, es aquel en el que el número de sucesos elementales que lo conforman es finito.

**1.17 Definición.** *Espacio muestral numerable* asociado a un experimento aleatorio, es aquel en el que se puede establecer una aplicación biyectiva entre los sucesos elementales que lo conforman y el conjunto de los números naturales  $\mathbb{N}$ .

**1.18 Definición.** *Espacio muestral no numerable* asociado a un experimento aleatorio, es aquel en el que no se puede establecer una correspondencia biunívoca entre los sucesos elementales que lo conforman y el conjunto  $\mathbb{N}$ .

Si  $Q$  es álgebra de Boole cuyos elementos son subconjuntos del espacio muestral  $\Omega$ , por la definición 1.4, tenemos que  $Q = 2^\Omega$ .

Extendamos la definición de probabilidad a espacios muestrales numerables (no necesariamente finitos). Sean  $\Omega$  el espacio muestral numerable y  $Q$  álgebra de Boole de subconjuntos de  $\Omega$ . Definamos una medida de probabilidad  $P$  en  $Q$  exigiendo la aditividad numerable además de la aditividad finita. Esto es, para toda colección numerable  $\{A_1, A_2, \dots\}$  de elementos de  $Q$ , exigimos que

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \quad \text{donde } A_i \cap A_j = \emptyset \text{ e } i \neq j \quad (1.9)$$

Las funciones de conjunto de aditividad finita que satisfacen la ecuación (1.9) se llaman de *aditividad numerable* o *completamente aditivas*. Naturalmente, esta propiedad exige también suponer que la reunión numerable  $A_1 \cup A_2 \cup A_3 \dots$  pertenece a  $Q$  cuando cada  $A_k$  pertenezca a  $Q$ . No todas las álgebras de Boole tienen esta propiedad. Las que la tienen se llaman  $\sigma$ -álgebras de Boole. Un ejemplo es el álgebra booleana de todos los subconjuntos de  $\Omega$ .

**1.19 Definición.** Sea  $\Omega$  un espacio muestral *numerable* y  $Q$  una  $\sigma$ -álgebra de Boole cuyos elementos son subconjuntos de  $\Omega$ . Probabilidad de un evento  $A_i \in Q$ , es la función  $P$  que satisface los axiomas:

1.  $0 \leq P(A_i) \leq 1$ .
2.  $P(\Omega) = 1$ .

3.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$  donde  $B \in Q$ .
4.  $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{k=1}^{\infty} P(A_k)$ , donde  $A_i \cap A_j = \emptyset$  e  $i \neq j$ .

Cuando  $Q$  es el álgebra de Boole de todos los subconjuntos de  $\Omega$ , una función de probabilidad queda completamente determinada mediante sus valores para los subconjuntos de un sólo elemento, tales valores se llaman *probabilidades puntuales*. En este caso, todo subconjunto  $A$  de  $\Omega$  es finito o infinito numerable, y la probabilidad de  $A$  se calcula sumando las probabilidades puntuales para todos los elementos de  $A$ ,

$$P(A) = \sum_{x \in A} P(x)$$

La suma del segundo miembro tiene un número finito de sumandos o es una serie absolutamente convergente.

Hasta el momento hemos considerado un número numerable, ya sea finito o infinito, de sucesos que conforman al espacio muestral. Ahora extendamos la definición de probabilidad incluyendo los espacios muestrales no numerables.

Si siguiéramos el mismo proceso que para espacios muestrales numerables tendríamos que partir de un espacio muestral no numerable  $\Omega$  cualquiera y una  $\sigma$ -álgebra de Boole  $Q$  de subconjuntos de  $\Omega$  y definir una medida de probabilidad que fuera una función de conjunto  $P$  no negativa, de aditividad finita y de aditividad numerable definida sobre  $Q$  siendo  $P(\Omega) = 1$ . Esto origina ciertas dificultades técnicas, en las cuales no ahondaremos, que no se presentan cuando  $\Omega$  es numerable. Cuando el espacio muestral  $\Omega$  es finito o infinito numerable, todos los subconjuntos de  $\Omega$  se pueden considerar como eventos. Esto no ocurre cuando el espacio muestral es no numerable, pues es difícil considerar el conjunto formado por todos los subconjuntos posibles, existiendo subconjuntos que no pueden considerarse como eventos. Es por esta razón que es necesario imponer las siguientes restricciones iniciales al espacio muestral  $\Omega$  y al álgebra de Boole  $Q$ :

Primeramente, restringimos  $\Omega$  a ser un subconjunto del eje real  $\mathbb{R}$ , o del espacio  $\mathbb{R}^n$ . Para el álgebra booleana  $Q$  empleamos subconjuntos especiales de  $\Omega$  que, en el lenguaje de la moderna teoría de la integración, se llaman subconjuntos  *$\Omega$  medibles*.

**1.20 Definición.** Consideremos subconjuntos de  $\mathbb{R}$ . *Subconjuntos medibles* son aquellos que tienen las propiedades siguientes:

- Si  $\{A_1, A_2, A_3, \dots\}$  es una colección numerable de conjuntos medibles, la reunión  $A_1 \cup A_2 \cup A_3 \cup \dots$  también es medible.
- Si  $A$  es medible, asimismo lo es  $\mathbb{R} - A$ , complemento de  $A$ .
- Todo intervalo  $A$  (abierto, cerrado, semiabierto, finito o infinito) es medible.

Así, los conjuntos medibles de  $\mathbb{R}$  forman una  $\sigma$ -álgebra booleana que contiene los intervalos. Existe una  $\sigma$ -álgebra booleana mínima (en el sentido de la inclusión) que tiene esta propiedad; sus elementos se llaman *conjuntos de Borel* en atención al matemático francés Émile Borel (1871-1956), creador de la teoría de la medida. Análogamente, en un espacio de dimensión 2 existe un  $\sigma$ -álgebra booleana mínima que contiene todos los productos cartesianos de pares de intervalos; sus elementos son conjuntos de Borel. Los conjuntos de Borel en un espacio de dimensión  $n$  se definen de forma análoga.

De ahora en adelante, siempre que utilicemos un conjunto  $\Omega$  de números reales como espacio muestral, supondremos que este conjunto es un conjunto de Borel. Los subconjuntos de Borel de  $\Omega$  forman asimismo una  $\sigma$ -álgebra de Boole; supondremos que nuestras medidas de probabilidad están definidas sobre esas álgebras de Boole. Dichas álgebras son lo bastante amplias para incluir todos los eventos que se presentan en las aplicaciones corrientes de la teoría de probabilidades.

**1.21 Definición.** Sean  $\Omega$  el espacio muestral subconjunto de  $\mathbb{R}^n$  y  $Q$  la  $\sigma$ -álgebra de Boole de subconjuntos de Borel de  $\Omega$ . Una función de conjunto no negativa, completamente aditiva  $P$  definida en  $Q$  con  $P(\Omega) = 1$  se llama *medida de probabilidad*.

## 1.4. Álgebra de variables aleatorias

Definir qué es una variable aleatoria no es tarea fácil. Comencemos con argumentos heurísticos:

Supongamos que tenemos un experimento  $F$  cuyos posibles resultados  $\zeta$  son varios objetos que corresponden a un cierto evento del espacio muestral  $\Omega$ . Ahora, a cada  $\zeta$  asignémosle de acuerdo a alguna regla el número

$$\chi(\zeta)$$

A la relación entre los elementos  $\zeta$  del conjunto  $\Omega$  y ciertos números se le conoce como variable aleatoria y se denota como  $\chi$ .

Dicho de otro modo, la variable aleatoria  $\chi$  es una función cuyo dominio es el conjunto  $\Omega$  de todos los posibles resultados de un experimento.

Los posibles resultados de un experimento pueden ser objetos identificados durante el transcurso del experimento. Por ejemplo, águila, as de espadas, blanco, etc. . .

En otros experimentos los resultados pueden ser números reales: tiempo de espera, siete en una ruleta, etc. . . Hay que recalcar que el carácter numérico de este tipo de resultados, es sólo una manera más de identificarlos.

De manera que  $\zeta$  no es un número, sino un elemento del conjunto  $\Omega$ .

La regla de asociación de una variable aleatoria asigna un y sólo un número real a cada elemento del espacio muestral  $\Omega$ .

Y el rango de la variable aleatoria es el espacio muestral de números definidos por la regla de asociación; en resumidas cuentas, un conjunto de números.

*1.1 Ejemplo.* Supongamos que un experimento consiste en lanzar una moneda (legal) al aire en dos ocasiones. Si la variable aleatoria es el número de “águilas” obtenidas en los dos lanzamientos, entonces,

- El espacio muestral de este experimento es  $\Omega = \{aa, as, sa, ss\}$ .
- La regla de asociación es contar el número de águilas obtenidas después de los dos lanzamientos.
- El rango del experimento es  $R = \{0, 1, 2\}$ .

*1.2 Ejemplo.* Supongamos ahora que el experimento consiste en tirar un dado dos veces. Si la variable aleatoria es el número de puntos acumulados en los dos lanzamientos, entonces,

- El espacio muestral de este experimento es  
 $\Omega = \{(1, 1), (1, 2), (1, 3), \dots, (6, 4), (6, 5), (6, 6)\}$ .
  
- La regla de asociación es contar el número total de puntos obtenidos después de los dos lanzamientos.
  
- El rango del experimento es  
 $R = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ .

Algunos autores, como Tucker 1973 [36], afirman que el nombre de variable aleatoria es engañoso puesto que no se trata de una variable sino de una función.

Julius W. R. Dedekind (1831-1916) y Wilhelm Weber (1804-1891) por primera vez en 1882 tuvieron la idea de considerar en  $f(x) \in \mathbb{R}$  a  $f$  como variable y a  $x$  fija, al contrario de la tradición clásica. De acuerdo con la teoría clásica, la variable aleatoria es una función que asigna un valor numérico a cada suceso del espacio muestral. El enfoque de Dedekind y Weber es el inverso: un suceso fijo del espacio muestral asigna un valor numérico a cada variable aleatoria, es decir, la variable aleatoria es en efecto variable.

**1.22 Definición.** Sean  $X$  y  $Y$  variables aleatorias, donde  $X, Y \in F$  y  $s \in \Omega$ . *Álgebra  $F$  de variables aleatorias* se define con las siguientes operaciones:



Enfoque clásico	Enfoque de Dedekind y Weber
$(X + Y)(s) = X(s) + Y(s)$	$s(X + Y) = s(X) + s(Y)$
$(X - Y)(s) = X(s) - Y(s)$	$s(X - Y) = s(X) - s(Y)$
$(X \cdot Y)(s) = X(s) \cdot Y(s)$	$s(X \cdot Y) = s(X) \cdot s(Y)$
$\left(\frac{X}{Y}\right)(s) = \frac{X(s)}{Y(s)}, Y(s) \neq 0$	$s\left(\frac{X}{Y}\right) = \frac{s(X)}{s(Y)}, s(Y) \neq 0$
$(\gamma X)(s) = \gamma \cdot X(s)$	$s(\gamma X) = \gamma \cdot s(X)$

donde  $\gamma$  es una constante.

Esto quiere decir que:

- La suma de variables aleatorias es variable aleatoria.
- La resta de variables aleatorias es variable aleatoria.
- El producto de variables aleatorias es variable aleatoria.
- El cociente de variables aleatorias es variable aleatoria.
- Una constante multiplicada por una variable aleatoria es variable aleatoria.

### 1.4.1. Relación entre probabilidad y álgebra de variables aleatorias

Ha llegado pues el momento de definir formalmente el concepto de variable aleatoria. Papoulis 1965 [26] propone la siguiente definición:

**1.23 Definición.** Variable aleatoria  $\chi$  es una función cuyo dominio es el espacio muestral  $\Omega$ , (i.e., el proceso de asignarle un número  $\chi(\zeta)$  a cada resultado  $\zeta$  del experimento  $F$ ) tal que:

1. El conjunto  $\{\chi \leq x\}$  es un suceso para cualquier número real  $x$ .
2. La probabilidad de los sucesos  $\{\chi = +\infty\}$  y  $\{\chi = -\infty\}$  es igual a cero:

$$P[\chi = +\infty] = P[\chi = -\infty] = 0$$

El tipo más simple de variable aleatoria lo constituye el indicador  $I$ .

**1.24 Definición.** Sean  $\mathcal{Q} = 2^\Omega$  álgebra de Boole y  $\Omega$  el espacio muestral, donde  $A \in \mathcal{Q}$  y  $s \in \Omega$ . El *indicador* de  $A$ ,  $I_A$  se define como la función:

$$I_A : \Omega \rightarrow \mathbb{N}$$

Esto es:

$$I_A(s) = \begin{cases} 0 & \text{si } s \notin A \\ 1 & \text{si } s \in A \end{cases}$$

**1.5 Teorema.** *El indicador es idempotente.*

*Demostración.* Por álgebra de variables aleatorias

$$(I_A \cdot I_A)(s) = I_A(s) \cdot I_A(s) = I_A(s)$$

es decir,  $I_A(s)$  es idempotente.  $\square$

**1.6 Teorema.** *Sean  $A, B \subset \Omega$  e  $I_A, I_B \in F$ . Entonces  $I_A(s) \cdot I_B(s) = I_{A \cap B}(s)$ .*

*Demostración.* Por álgebra de variables aleatorias  $(I_A \cdot I_B)(s) = I_A(s) \cdot I_B(s)$  y

$$\begin{aligned} I_A(s) \cdot I_B(s) &= \begin{cases} 0 & \text{si } s \notin A \cap B \\ 1 & \text{si } s \in A \cap B \end{cases} \\ &= I_{A \cap B}(s) \end{aligned}$$

$\square$

**1.7 Teorema.** *Sean  $A, B \subset \Omega$  e  $I_A, I_B \in F$ . Entonces*

$$I_A(s) + I_B(s) - I_A(s) \cdot I_B(s) = I_{A \cup B}(s)$$

**1.8 Teorema.** Sean  $X, Y \in F$  variables aleatorias idempotentes. Entonces  $(X \cap Y)^2 = X \cap Y$ , es decir,  $X \cap Y$  es idempotente.

*Demostración.* Por el teorema 1.6,  $X \cap Y = X \cdot Y$  y entonces

$$\begin{aligned} (X \cap Y)^2 &= (X \cdot Y)^2 \\ &= X \cdot Y \cdot X \cdot Y \\ &= X \cdot X \cdot Y \cdot Y \\ &= X \cdot Y \\ &= X \cap Y \end{aligned}$$

□

**1.9 Teorema.** Sean  $X, Y \in F$  variables aleatorias idempotentes. Entonces  $(X \cup Y)^2 = X \cup Y$ , es decir,  $X \cup Y$  es idempotente.

*Demostración.* Por el teorema 1.7,  $X \cup Y = X + Y - X \cdot Y$  y entonces

$$\begin{aligned} (X \cup Y)^2 &= (X + Y - X \cdot Y)^2 \\ &= (X + Y - X \cdot Y)(X + Y - X \cdot Y) \\ &= X + Y - X \cdot Y \\ &= X \cup Y \end{aligned}$$

□

Tras enunciar estos teoremas podemos concluir que el álgebra de variables aleatorias satisface las leyes del álgebra de Boole y que, por lo tanto, los conceptos de probabilidad y variable aleatoria se desprenden de la definición del álgebra booleana.

La Figura 1.1 muestra que los conceptos de probabilidad y de álgebra de variables aleatorias son dos mundos separados: probabilidad es una función que asigna valores entre cero y uno a cada evento del álgebra de Boole  $Q$  mientras que álgebra de variables aleatorias es una función que asigna un valor

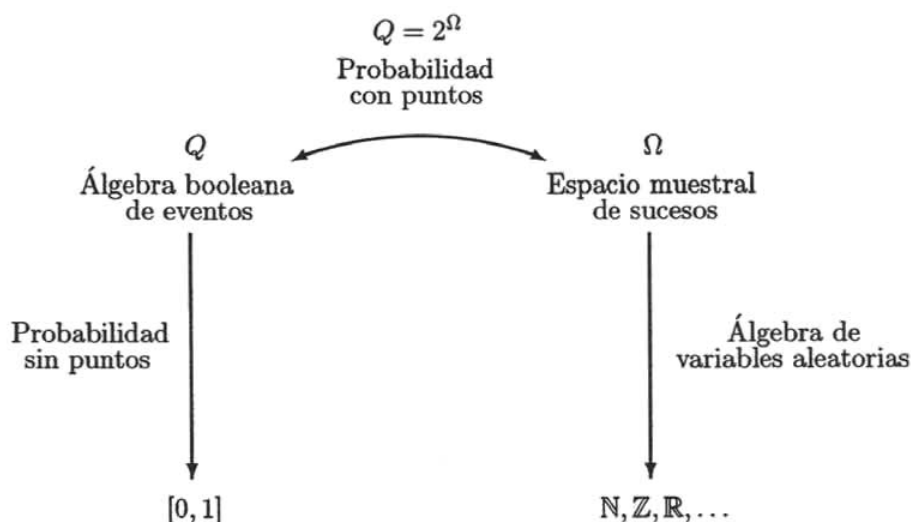


Figura 1.1: Probabilidad y álgebra de variables aleatorias

numérico a cada suceso del espacio muestral  $\Omega$ . La relación entre estos dos mundos se establece si  $Q = 2^\Omega$ , es decir, si  $Q$  es álgebra de Boole de todos los subconjuntos del espacio muestral  $\Omega$ , en cuyo caso la función de probabilidad queda completamente determinada mediante sus valores para los subconjuntos de un sólo elemento, por lo que recibe el nombre de *probabilidad con puntos* o *probabilidad puntual*.

Sin embargo, cabe destacar que cuando no se establece esta relación y no se toma en cuenta el espacio muestral  $\Omega$  sólo se habla del concepto de *probabilidad abstracta* o *probabilidad sin puntos*.

### 1.4.2. Variables aleatorias discretas

Se dice que una variable aleatoria  $\chi$  es *discreta* si el conjunto de posibles valores de  $\chi$  es numerable (pero no necesariamente finito). Dado que no conocemos con certeza el valor que  $\chi$  tendrá, decimos que tomará un valor  $x$  con probabilidad  $p_\chi(x)$ . Esto es

$$p_\chi(x) = P[\chi = x] \quad (1.10)$$

A lo cual se le conoce como *función masa de probabilidad* de  $\chi$ . Como una distribución de probabilidad discreta es una función probabilística, obedece las reglas y propiedades de la teoría de la probabilidad. Esto significa, entre otras cosas, que  $0 \leq p_\chi(x) \leq 1 \forall x$  y que

$$\sum_x p_\chi(x) = 1$$

Nótese además que  $x$  puede ser cualquier número real.

Si se desea saber la probabilidad con la que  $\chi$  tomará valores  $x_i$  menores o iguales a algún valor  $a$ , el siguiente concepto es de gran utilidad.

**1.25 Definición.** La *función de distribución acumulativa* de una variable aleatoria  $\chi$  es la función

$$F_\chi(a) = P[\chi \leq a] = \sum_{x \leq a} p_\chi(x)$$

definida para toda variable real  $a$ .

Obviamente,  $0 \leq F_\chi(x) \leq 1$ . Si partimos del hecho de que

$$\{\chi \leq b\} = \{\chi \leq a\} \cup \{a < \chi \leq b\}$$

entonces

$$F_\chi(b) = F_\chi(a) + P[a < \chi \leq b]$$

de donde se sigue que

$$P[a < \chi \leq b] = F_\chi(b) - F_\chi(a) \tag{1.11}$$

### 1.4.3. Variables aleatorias continuas

Se dice que una variable aleatoria  $\chi$  es *continua* si existe una función no negativa  $f_\chi(x)$  tal que la función de distribución acumulativa  $F_\chi(x)$  puede calcularse como

$$F_\chi(a) = P[\chi \leq a] = \int_{-\infty}^a f_\chi(x) dx \tag{1.12}$$

La función  $f_\chi(x)$  es la *función densidad de probabilidad* de la variable aleatoria  $\chi$ . Lógicamente, dado que hablamos de probabilidades debe satisfacerse la ecuación

$$\int_{-\infty}^{\infty} f_\chi(x) dx = 1$$

Análogo a (1.11), la probabilidad de que la variable aleatoria  $\chi$  se encuentre en el intervalo  $(a, b)$  se calcula

$$P[a < \chi \leq b] = \int_a^b f_\chi(x) dx \quad (1.13)$$

La función de densidad  $f_\chi(x)$  no necesariamente es una función continua. Sin embargo, la función de distribución acumulativa  $F_\chi(x)$  automáticamente lo es. Lo cual implica

$$P[\chi = x] = 0 \quad (1.14)$$

para cualquier valor de  $x$ . Esto quiere decir que los eventos

$$\{a \leq \chi < b\} \quad \{a < \chi \leq b\} \quad \{a < \chi < b\}$$

tienen la misma probabilidad dada por (1.13).

Es claro que para calcular la función de densidad a partir de la función de distribución acumulativa

$$f_\chi(x) = \frac{d}{dx} F_\chi(x) \quad (1.15)$$

#### 1.4.4. Momentos de una variable aleatoria

En muchos casos estamos interesados en conocer algunos valores característicos de una función de distribución más que en la función de distribución misma. Entre los valores característicos que más comúnmente se desean conocer se encuentra el *valor promedio* de una variable aleatoria *positiva*  $\chi(t)$ , es decir, su *valor esperado* denotado por  $E[\chi]$  o bien por  $\bar{\chi}$  cuyo valor se define como

$$E[\chi] = \int_0^{\infty} t f_\chi(t) dt$$

Esta ecuación es el *primer momento*. Es importante observar que estamos definiendo la integral sólo en el intervalo  $t \in [0, \infty)$  dado que la variable independiente siempre será el tiempo para nuestro estudio particular sobre procesos estocásticos.

A veces, también es necesario conocer el valor esperado de la  $n$ -ésima potencia de una variable aleatoria, es decir, el  *$n$ -ésimo momento*. El  $n$ -ésimo momento puede calcularse a través de la fórmula

$$E[\chi^n] = \int_0^{\infty} t^n f_{\chi}(t) dt$$

Y el  *$n$ -ésimo momento central* de una variable aleatoria se define como

$$\overline{(\chi - \bar{\chi})^n} = \int_0^{\infty} (t - \bar{\chi})^n f_{\chi}(t) dt$$

La *varianza* es el *segundo momento central* y se denota como  $\sigma_{\chi}^2$  y, utilizando la expresión anterior, se calcula

$$\begin{aligned} \sigma_{\chi}^2 &= \overline{(\chi - \bar{\chi})^2} \\ &= \overline{\chi^2} - (\bar{\chi})^2 \end{aligned}$$

Por último, a la raíz cuadrada  $\sigma_{\chi}$  de la varianza se le llama *desviación estándar*. Al radio de la desviación estándar a la media de una variable aleatoria se le llama *coeficiente de variación* y se calcula

$$C_{\chi} = \frac{\sigma_{\chi}}{\bar{\chi}}$$

### 1.4.5. Distribución conjunta de variables aleatorias

Utilicemos el símbolo  $\mathbb{R}^n$  para denotar el conjunto de todas las  $n$ -tuplas de números reales. Sean  $\chi_1, \chi_2, \dots, \chi_n$  variables aleatorias. Se dice que estas variables aleatorias tienen una *distribución conjunta discreta* si existe una función no negativa  $p(x_1, x_2, \dots, x_n)$  de  $n$  variables reales que toma el valor de 0 excepto en un conjunto numerable de puntos en  $\mathbb{R}^n$ , tal que

$$P[\chi_1 = x_1, \chi_2 = x_2, \dots, \chi_n = x_n] = p(x_1, x_2, \dots, x_n)$$

para todos los puntos  $x_1, x_2, \dots, x_n$  en  $\mathbb{R}^n$ . Obviamente, tenemos que

$$\sum_{x \in \mathbb{R}^n} p(x) = 1$$

La colección de variables aleatorias  $\chi_1, \chi_2, \dots, \chi_n$  tiene una *distribución conjunta continua* si existe una función no negativa integrable  $f(x_1, x_2, \dots, x_n)$  de  $n$  variables reales que satisface

$$P[\chi_1 \leq a_1, \dots, \chi_n \leq a_n] = \int_{-\infty}^{a_1} \dots \int_{-\infty}^{a_n} f(x_1, x_2, \dots, x_n) dx_1 \dots dx_n$$

para todo límite superior  $a_1, a_2, \dots, a_n$ . La función  $f$  es la *función de densidad de probabilidad conjunta* de las variables aleatorias  $\chi_1, \chi_2, \dots, \chi_n$  y, como en el caso discreto, debemos tener que

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_1 \dots dx_n = 1$$

Si conocemos la distribución conjunta  $f$  de  $\chi_1, \chi_2, \dots, \chi_n$ , podemos obtener la distribución de cualquiera de estas variables, digamos  $\chi_m$ , integrando sobre todos los valores de las variables aleatorias restantes. Esto es  $f_m(x) =$

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_{m-1}, x, x_{m+1}, \dots, x_n) dx_1 \dots dx_{m-1} dx_{m+1} \dots dx_n$$

es la función densidad de probabilidad de  $\chi_m$ . Esto también se cumple para el caso discreto en el que sumaremos, en lugar de integrar, sobre todos los valores de las demás variables.

## 1.5. Proceso estocástico

La palabra estocástico proviene del griego  $\sigma\tau\omicron\chi\alpha\sigma\tau\iota\kappa\omicron\varsigma$ , que en el siglo XVII significaba hábil en conjeturar. Sin embargo y aunque no se sabe cómo, poco a poco adquirió el significado de perteneciente o relativo al azar, significado que prevalece a nuestros días.

**1.26 Definición.** *Proceso* es una sucesión de estados.



En los libros de procesos estocásticos de Coleman 1976 [10], Parzen 1962 [27] y Ross 1996 [30], entre otros, se establece la siguiente definición de *proceso estocástico*.

**1.27 Definición.** *Proceso estocástico* es la familia de variables aleatorias  $\{\chi_t, t \in T\}$ , donde  $T$  es llamado espacio parametral de tiempo y donde para cada  $t \in T$ ,  $\chi_t$  es un punto en el espacio  $S$ , llamado espacio de estados.

Sin embargo, Vélez 1996 [37] precisa que proceso estocástico es una *sucesión de aplicaciones medibles*

$$\chi_n : (\Omega, \mathcal{Q} = 2^\Omega) \rightarrow (\mathbb{R}, \mathbb{B})$$

donde  $\mathbb{B}$  representa la  $\sigma$ -álgebra de Borel en la recta real.

Cabe señalar que la relación entre los conceptos de estado, suceso y variable aleatoria no se ha establecido formalmente, sino de manera intuitiva, ya que se asume que el espacio de estados  $S$  equivale al espacio muestral  $\Omega$ .

La clasificación de los procesos estocásticos, también conocidos como procesos aleatorios, depende de tres factores: el *espacio de estados*, la naturaleza del *espacio parametral (o paramétrico) del tiempo* y de la *dependencia estadística* entre las variables aleatorias  $\chi_t$  para diferentes valores del espacio parametral del tiempo.

**1.28 Definición.** Si los valores en el espacio de estados  $S$  de  $\chi_t$  son finitos o numerables decimos que el proceso es de estados discretos, a lo cual comúnmente se le llama *cadena*. El espacio de estados de una cadena usualmente es el conjunto de enteros  $\{0, 1, 2, \dots\}$ . En cambio, si los valores permitidos en el espacio de estados están definidos sobre un intervalo continuo ya sea finito o infinito, entonces decimos que el proceso es de estados continuos.

La teoría de los procesos estocásticos con espacio de estados continuo es sumamente complicada y nosotros nos limitaremos a considerar únicamente el caso de los procesos estocásticos con espacio de estados discreto.

**1.29 Definición.** Si los valores de  $t$  son finitos o numerables, entonces decimos que tenemos un proceso de tiempo discreto. Y si  $t$  pertenece a un intervalo de tiempo continuo finito o infinito, se dice que el proceso es de tiempo continuo.

Cuando  $t$  sea discreto escribiremos  $\chi_n$  en lugar de  $\chi_t$  para referirnos a una *secuencia estocástica* en lugar de a un proceso estocástico.

**1.30 Definición.** Considere la distribución conjunta de todas las variables aleatorias  $\mathbf{X} = \{\chi_{t_1}, \chi_{t_2}, \dots\}$  dada por

$$F_{\mathbf{X}}(\mathbf{x}; \mathbf{t}) = P[\chi_{t_1} \leq x_1, \dots, \chi_{t_n} \leq x_n]$$

para toda  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ ,  $\mathbf{t} = (t_1, t_2, \dots, t_n)$  y toda  $n$ . Entonces la naturaleza de  $F_{\mathbf{X}}(\mathbf{x}; \mathbf{t})$  es la tercera cantidad que determina la clase del proceso estocástico.

En el siguiente capítulo abordaremos únicamente a los procesos estocásticos conocidos como cadenas markovianas y haremos una breve introducción a los procesos markovianos.



# Capítulo 2

## Teoría markoviana

*“Las Matemáticas no son una marcha cautelosa a lo largo de una carretera bien despejada, sino un viaje por un desierto desconocido en el que los exploradores se pierden a menudo”.*

W.S. Anglin

En este capítulo se hace un estudio extenso de los procesos estocásticos conocidos como cadenas de Markov, bajo un enfoque práctico a través de varios ejemplos que ilustran el concepto que involucra cada una de las definiciones matemáticas formales. Asimismo, se hace una breve introducción a los procesos markovianos.

### 2.1. Cadenas de Markov

El matemático Andrei Andreyevich Markov (1856-1922) comenzó su carrera docente en la Universidad de San Petesburgo, en su natal Rusia, en el año 1886. Aunque realizó numerosos estudios sobre teoría de números, fracciones continuas, límites de integrales, teoría de la aproximación y convergencia de series, Markov es recordado principalmente por sus estudios sobre secuencias de variables aleatorias en las cuales la variable futura es determinada única-

mente por la variable presente, es decir, esta es dependiente sólo de la variable antecesora inmediata. De ahí que dicha secuencia de variables aleatorias tome el nombre de *cadena de Markov*.

La importancia de las *cadena de Markov* se resume principalmente en dos puntos:

- (i) Existe un gran número de fenómenos físicos, biológicos, económicos y sociales que pueden ser descritos a través de ellas.
- (ii) Existe además una teoría desarrollada que nos permite realizar los cálculos necesarios para el estudio de dichos fenómenos.

**2.1 Definición.** La secuencia estocástica  $\{\chi_n | n = 0, 1, 2, \dots\}$ , es decir, con espacio parametral de tiempo  $T \in \mathbb{N}$  y un espacio de estados  $S$  finito o infinito numerable, es una *cadena de Markov* si se cumple que

$$P[\chi_{n+1} = j | \chi_0 = i_0, \chi_1 = i_1, \dots, \chi_n = i] = P[\chi_{n+1} = j | \chi_n = i] \quad (2.1)$$

Cuando el espacio de estados  $S$  es finito, la cadena de Markov es *finita*.

La expresión

$$p_{ij} = P[\chi_{n+1} = j | \chi_n = i]$$

es la *probabilidad de transición en un paso* y denota la probabilidad de que la cadena vaya del estado  $i$  al estado  $j$  en un sólo paso. Es importante señalar que  $p_{ij} = p_{ij}^1$ .

La forma general para expresar la *probabilidad de transición en  $m$  pasos* es

$$p_{ij}^m = P[\chi_{n+m} = j | \chi_n = i]$$

esto es, la probabilidad de que el sistema pase del estado  $i$  al  $j$  en  $m$  pasos.

Es importante enfatizar que las *probabilidades de transición en  $m$  pasos* se refieren a *todos* los posibles caminos para llegar de un estado a otro. Nótese además que las probabilidades  $p_{ij}^m$  deben satisfacer:

$$0 < p_{ij}^m \leq 1, \quad i, j = 1, 2, \dots, N; \quad m = 0, 1, 2, \dots$$

$$\sum_{j \in S} p_{ij}^m = 1, \quad i = 1, 2, \dots, N; \quad m = 0, 1, 2, \dots$$

**2.2 Definición.** Se dice que una cadena de Markov es *homógena* en el tiempo si la siguiente condición se cumple,

$$P[\chi_{n+1} = j | \chi_n = i] = P[\chi_{m+1} = j | \chi_m = i] = p_{ij} \quad \text{con } m \neq n$$

es decir, si todas las probabilidades de transición  $p_{ij}$  son fijas e independientes del tiempo.

Para nuestro estudio consideraremos únicamente cadenas markovianas homogéneas.

Supongamos que la cadena de Markov llega al estado  $i$  al tiempo 0, y que el proceso no deja al estado  $i$  (lo cual quiere decir que no ocurre transición a otro estado) durante las siguientes  $s$  unidades de tiempo. ¿Cuál es la probabilidad de que la cadena no deje al estado  $i$  durante las siguientes  $t$  unidades de tiempo? Para responder a esto, note que del hecho de que la cadena se encuentre en el estado  $i$  al tiempo  $s$  se sigue, por la propiedad markoviana, que la probabilidad de que se quede en dicho estado en el intervalo de tiempo  $[s, s + t]$ , es la probabilidad de que permanezca en el estado  $i$  por al menos  $t$  unidades de tiempo, lo cual quiere decir que si la variable aleatoria  $N_i$  denota el tiempo que la cadena permanece en el estado  $i$  antes de hacer una transición a cualquier otro estado, entonces no tiene memoria y la única distribución de probabilidad discreta que satisface dicha condición es la geométrica. Esta distribución tiene un sólo parámetro, llámese  $p$ , donde la media es  $\frac{(1-p)}{p}$  y la función masa de probabilidad es

$$p_{N_i}(n) = P[N_i = n] = (1-p)^{(n-1)}p, \quad \text{para } n = 1, 2, \dots; 0 \leq p \leq 1$$

y calculando su función de distribución acumulativa

$$F_{N_i}(t) = P[N_i \leq t] = \sum_{n \leq t} p_{N_i}(n)$$

obtendremos la probabilidad de que la cadena no deje al estado  $i$  durante las siguientes  $t$  unidades de tiempo.

Supongamos ahora que nuestra cadena se encuentra en el estado  $i \in S$  al tiempo  $r \in \mathbb{N}$ . La probabilidad de ir del estado  $i$  al estado  $j$  en un número de pasos determinado, digamos  $n + m$ , equivale a la probabilidad de partir del

estado  $i$  y llegar a un estado intermedio  $k$  al tiempo  $n$  y de ahí llegar al estado  $j$  en  $m$  pasos. Los eventos  $\{\chi_{r+n} = k | \chi_r = i\}$  y  $\{\chi_{r+n+m} = j | \chi_{r+n} = k\}$  son independientes y utilizando la propiedad markoviana tenemos

$$P[\chi_{r+n+m} = j | \chi_{r+n} = k, \chi_r = i] = P[\chi_{r+n+m} = j | \chi_{r+n} = k]$$

de manera que podemos escribir la siguiente ecuación recursiva considerando todos los posibles estados intermedios  $k$

$$p_{ij}^{n+m} = \sum_{k \in S} P[\chi_{r+n} = k | \chi_r = i] P[\chi_{r+n+m} = j | \chi_{r+n} = k]$$

Esto es

$$p_{ij}^{n+m} = \sum_{k \in S} p_{ik}^n p_{kj}^m \quad (2.2)$$

A la ecuación (2.2) se le conoce como la *ecuación de Chapman-Kolmogorov* para cadenas de Markov de tiempo discreto.

Podemos escribir las *probabilidades de transición en un paso* de forma matricial

$$\begin{array}{c} \chi_{n+1} \\ 0 \quad 1 \quad 2 \quad \dots \\ \chi_n \end{array} \begin{pmatrix} p_{00} & p_{01} & p_{02} & \dots \\ p_{10} & p_{11} & p_{12} & \dots \\ p_{20} & p_{21} & p_{22} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} = \mathbf{P}$$

La *matriz de probabilidades de transición*  $\mathbf{P}$  es *estocástica* si cada uno de sus renglones suma uno y *doblemente estocástica*<sup>1</sup> si tanto sus renglones como sus columnas suman uno. Utilizando la ecuación de *Chapman-Kolmogorov*, es fácil verificar que

$$\begin{aligned} \mathbf{P}^2 &= \mathbf{P}\mathbf{P} \\ \mathbf{P}^3 &= \mathbf{P}^2\mathbf{P} \\ \mathbf{P}^4 &= \mathbf{P}^3\mathbf{P} \\ &\vdots \end{aligned}$$

<sup>1</sup>O bien estocástica por duplicado.

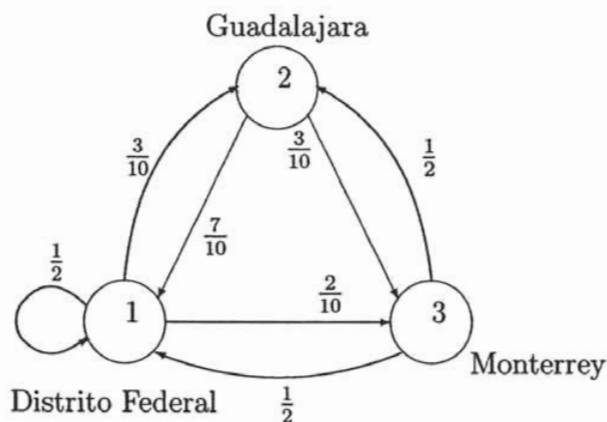


Figura 2.1: Cadena de Markov ergódica

Y en general,

$$\mathbf{P}^m = \mathbf{P}^{m-1}\mathbf{P} \quad (2.3)$$

en donde

$$\mathbf{P}^0 = \mathbf{I}$$

la matriz identidad.

Esto nos permite calcular las probabilidades de transición en  $m$  pasos a partir de la  $n$ -ésima potencia de la matriz de transición en un paso.

Para estudiar el comportamiento de las cadenas de Markov de una forma más sencilla, la teoría de grafos resulta de gran utilidad. Los estados suelen denotarse por nodos, mientras que los arcos muestran las transiciones posibles de un estado a otro y generalmente van acompañados de sus respectivas probabilidades. Al diagrama resultante se le llama *diagrama de transición de estados*.

*2.1 Ejemplo.* El señor Homobono Buenrostro es agente de ventas de una compañía productora de equipo de seguridad industrial. Por ser el mejor vendedor, al buen Homobono le han encomendado la venta de calzado de seguridad industrial en el Distrito Federal, Guadalajara y Monterrey.

Cada semana, digamos cada domingo, Homobono debe decidir viajar a alguno de esos tres destinos para levantar pedidos.



Las preferencias de Homobono, quien conoce el mercado, son las siguientes:

Si se encuentra en el Distrito Federal, viajará a Guadalajara con probabilidad 0.3 o a Monterrey con probabilidad 0.2. Si está en Guadalajara, viajará al Distrito Federal con probabilidad 0.7 o a Monterrey con probabilidad 0.3. Y si se encuentra en Monterrey, viajará a cualquiera de los otros dos destinos con probabilidad 0.5.

Sea  $\chi_{n+1}$  el lugar en el que se encuentra el vendedor en la semana  $n + 1$ . Es claro que este ejemplo es una cadena de Markov, ya que el lugar en el que se encontrará Homobono está determinado por el lugar de procedencia  $\chi_n$ .

1. La Figura 2.1 muestra el comportamiento de nuestro amigo Homobono. Con nuestros destinos enumerados de este modo, el espacio de estados es  $S = \{1, 2, 3\}$  donde el estado 1 se refiere al Distrito Federal; el estado 2, a Guadalajara; y el estado 3, a Monterrey. La matriz de transición  $\mathbf{P}$  es la siguiente:

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} & 1 & 2 & 3 \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.7 & 0 & 0.3 \\ 0.5 & 0.5 & 0 \end{pmatrix} \end{array} \end{array}$$

2. La probabilidad de que Homobono llegue a Monterrey en dos semanas habiendo partido de Guadalajara puede calcularse a través de la segunda potencia de la matriz de transición  $\mathbf{P}$ :

$$\mathbf{P}^2 = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.7 & 0 & 0.3 \\ 0.5 & 0.5 & 0 \end{pmatrix} \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.7 & 0 & 0.3 \\ 0.5 & 0.5 & 0 \end{pmatrix} = \begin{pmatrix} 0.56 & 0.25 & 0.19 \\ 0.5 & 0.36 & 0.14 \\ 0.6 & 0.15 & 0.25 \end{pmatrix}$$

Entonces,  $p_{23}^2 = 0.14$

3. Utilizando la ecuación de Chapman-Kolmogorov podemos comprobar el resultado obtenido en 2:

$$\begin{aligned} p_{23}^2 &= \sum_{k=1}^3 p_{2k}^1 p_{k3}^1 \\ &= p_{21}p_{13} + p_{22}p_{23} + p_{23}p_{33} \\ &= (0.7)(0.2) + (0)(0.3) + (0.3)(0) = 0.14 \end{aligned}$$

4. La probabilidad de que el agente de ventas se encuentre en Guadalajara en la segunda semana  $y$  en Monterrey en la cuarta semana, siendo que en la primera semana se encontró en el Distrito Federal es la probabilidad conjunta de dos eventos condicionada a un evento:

$$P[\chi_4 = 3, \chi_2 = 2 | \chi_1 = 1] = \frac{P[\chi_4 = 3, \chi_2 = 2, \chi_1 = 1]}{P[\chi_1 = 1]}$$

Sabemos que  $P[\chi_4 = 3, \chi_2 = 2, \chi_1 = 1]$  es igual a

$$P[\chi_4 = 3 | \chi_2 = 2, \chi_1 = 1] P[\chi_2 = 2, \chi_1 = 1]$$

donde

$$P[\chi_2 = 2, \chi_1 = 1] = P[\chi_2 = 2 | \chi_1 = 1] P[\chi_1 = 1]$$

Y además por la propiedad markoviana

$$P[\chi_4 = 3 | \chi_2 = 2, \chi_1 = 1] = P[\chi_4 = 3 | \chi_2 = 2]$$

Entonces

$$\begin{aligned} P[\chi_4 = 3, \chi_2 = 2 | \chi_1 = 1] &= P[\chi_4 = 3 | \chi_2 = 2] P[\chi_2 = 2 | \chi_1 = 1] \\ &= p_{23}^2 p_{12} \\ &= (0.14)(0.3) = 0.042 \end{aligned}$$

Pero, ¿qué ocurre cuando el estado inicial es aleatorio? La *probabilidad de estado*  $\pi_j^m$  de encontrar a la cadena de Markov en el estado  $j$  en el  $m$ -ésimo paso:

$$\pi_j^{(m)} = P[\chi_m = j]$$

Sabemos

$$P[\chi_m = j | \chi_0 = i] = p_{ij}^m$$

donde hemos asumido, sin pérdida de generalidad, que estamos en el estado  $i$  al tiempo 0. Multiplicando por  $\pi_i^{(0)} = P[\chi_0 = i]$  a ambos miembros de la ecuación, sumando sobre todos los estados  $i$  y aplicando el teorema de probabilidad total, se tiene que

$$\begin{aligned} \sum_i P[\chi_0 = i] P[\chi_m = j | \chi_0 = i] &= \sum_i P[\chi_0 = i] p_{ij}^m \\ P[\chi_m = j] &= \sum_i \pi_i^{(0)} p_{ij}^m \end{aligned}$$

Es decir

$$\pi_j^{(m)} = \sum_i \pi_i^{(0)} p_{ij}^m \quad (2.4)$$

Lo cual permite calcular la distribución de  $\chi_m$  en términos de la distribución inicial  $\pi_i^{(0)}$  y la función de transición en  $m$  pasos.

El vector renglón conformado por las probabilidades de estado al tiempo  $m$  es el vector de probabilidades  $\Pi^{(m)}$ . Esto es

$$\Pi^{(m)} = (\pi_0^{(m)}, \pi_1^{(m)}, \pi_2^{(m)}, \dots)$$

Con esta definición, la ecuación (2.4) puede escribirse matricialmente

$$\Pi^{(m)} = \Pi^{(0)} \mathbf{P}^m, \quad m = 1, 2, \dots \quad (2.5)$$

*2.2 Ejemplo.* Siguiendo con el ejemplo del vendedor de la Figura 2.1

1. Para calcular la probabilidad conjunta  $P[\chi_2 = 3, \chi_1 = 1]$ , como no se ha especificado hasta este momento cuál es la probabilidad de estado inicial, el resultado queda expresado como:

$$\begin{aligned} P[\chi_2 = 3, \chi_1 = 1] &= P[\chi_2 = 3 | \chi_1 = 1] P[\chi_1 = 1] \\ &= p_{13} \pi_1^{(1)} \end{aligned}$$

2. Suponiendo que nuestra cadena comience en el estado 3, el vector de probabilidades de estado al tiempo 1:

$$\Pi^{(1)} = (0, 0, 1) \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.7 & 0 & 0.3 \\ 0.5 & 0.5 & 0 \end{pmatrix} = (0.5, 0.5, 0)$$

Como veremos posteriormente, el vector de probabilidades de estado  $\Pi^{(m)}$  tiende a un límite para  $m \rightarrow \infty$ . Puede mostrarse que para ciertas cadenas markovianas de tiempo discreto el efecto de  $\Pi^{(0)}$  sobre el vector  $\Pi^{(m)}$  se desvanece por completo. Este fenómeno no prevalece para todas las cadenas de Markov.

Para predecir el comportamiento de una cadena de Markov en el largo plazo, es decir, cuando el número de transiciones tiende a infinito, es necesario hacer un estudio minucioso de las características de los estados que la conforman.

Es por ello que ahora nos enfocaremos a realizar una clasificación de estados de acuerdo a las definiciones que siguen.

### 2.1.1. Clasificación de los estados

Consideremos los estados  $i, j \in S$ . Si existe un camino de  $i$  a  $j$ , es decir, si existe un entero  $n$  tal que

$$p_{ij}^n > 0$$

entonces escribimos  $i \rightarrow j$ .

Se dice que dos estados se *comunican* si existe un camino del estado  $i$  para llegar al estado  $j$  y viceversa y esto se denota  $i \rightleftharpoons j$ .

Sea  $C[i] = \{j | i \rightleftharpoons j; j \in S\} \forall i \in S$ . Llamamos  $C[i]$  a la *clase* del estado  $i$ .

**2.3 Definición.** Se dice que el conjunto de estados  $C$  de una cadena de Markov es *irreducible* si todos los estados en  $C$  se comunican entre sí, i.e.

$$p_{ij}^m > 0 \quad \forall m \geq 1, \quad i, j \in C$$

Si una cadena de Markov es irreducible, tendrá sólo una clase de estados, i.e.  $C[i] = C[j] \forall i, j \in S$ .

Sea  $C$  una clase cualquiera de estados y  $\overline{C}$  el conjunto de todos los estados de la cadena de Markov que no pertenecen a la clase  $C$ .

**2.4 Definición.** Se dice que una clase  $C$  es *cerrada* si las probabilidades de transición  $p_{ij}^m = 0 \quad \forall m \geq 1$  con  $i \in S$  y  $j \notin S$ , i.e., si no existe un estado en  $C$  que lleve a otro estado en  $\overline{C}$ .

Si una clase cerrada  $C$  está conformada únicamente por un estado, digamos  $i$ , entonces a  $i$  se le llama *estado absorbente*. Una condición suficiente y necesaria para que  $i$  sea un estado absorbente es que

$$p_{ii} = 1$$

lo cual evidentemente implica

$$p_{ij} = 0 \quad \forall i \neq j$$

esto es, un estado absorbente no se comunica con ningún otro estado.

Por otra parte, todos los estados de una cadena irreducible deben formar un conjunto cerrado y ningún subconjunto puede ser cerrado.

**2.5 Definición.** Se dice que una clase  $C$  es *transitoria* si existe algún camino que lleve fuera de  $C$ . Esto es, si  $\exists i \in C$  y  $k \in \bar{C}$  tal que  $p_{ik} > 0$ . Cada uno de los estados pertenecientes a una clase transitoria es transitorio.

**2.6 Definición.** Una cadena de Markov es *absorbente* si cada estado en ella es absorbente y/o transitorio.

**2.7 Definición.** Se dice que una clase  $C$  es *ergódica* si cada camino que comienza en  $C$  vuelve a  $C$ . Esto es

$$\sum_{j \in C} p_{ij} = 1, \quad \forall i \in C$$

A cada estado de una clase ergódica se le llama *ergódico*. Una cadena de Markov irreducible consiste en una clase ergódica particular, i.e.,  $C[i] = S, \forall i \in S$ .

Las definiciones 2.5, 2.6 y 2.7 nos hablan sobre las características de *transitoriedad*, *absorción* y *ergodicidad* de una clase.

Estos mismos conceptos serán retomados para el caso de los estados que conforman una clase.

**2.3 Ejemplo.** Considere la siguiente cadena de Markov

$$\mathbf{P} = \begin{array}{c} \begin{array}{cccc} & 0 & 1 & 2 & 3 \\ \begin{array}{l} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0.1 & 0.4 & 0.5 & 0 \\ 0 & 0 & 1 & 0 \\ 0.2 & 0 & 0.5 & 0.3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \end{array}$$

La figura 2.2 muestra que los cuatro estados *no* constituyen una cadena irreducible, pues los estados 0, 1 y 2 no pueden ser alcanzados desde el estado 3.

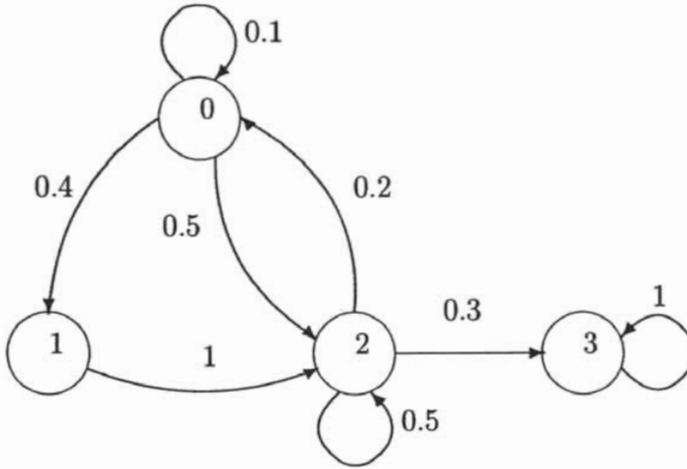


Figura 2.2: Cadena reducible

El estado 3 por sí mismo forma una cadena cerrada e irreducible y, como consta de un único elemento, es además absorbente.

Formalmente,  $C[3] = 3$ ,  $C[0] = \{0, 1, 2\}$

Los estados 0, 1 y 2 son transitorios, pues todos ellos pueden alcanzar al estado 3 y una vez que la cadena caiga en este estado, no habrá posibilidad de que la cadena vuelva a la clase  $C[0]$ .

**2.8 Definición.** Sea la v.a.  $T_{ij} = \min\{n \geq 1 : \chi_n = j | \chi_0 = i\}$ <sup>2</sup> el tiempo que tarda la cadena en llegar, por primera vez, al estado  $j$  habiendo partido de  $i$ . A la v.a.  $T_{ij}$  suele llamársele *tiempo de alcance* o *tiempo de primera pasada*<sup>3</sup>.

La *probabilidad de que el tiempo de primera pasada del estado  $i$  al estado  $j$  sea igual a  $n$*  se denota

$$f_{ij}^{(n)} = P\{\chi_n = j, \chi_r \neq j \forall r < n | \chi_0 = i\} = P_i\{T_j = n\}$$

<sup>2</sup>Note que encontrarse en el estado  $j$  al tiempo 0 no cuenta.

<sup>3</sup>Algunos autores prefieren llamar a  $T_{ij}$  *tiempo de llegada*, mientras que otros se refieren únicamente al caso en que  $i = j$  y utilizan la definición  $T_j = \min\{n \geq 1 : \chi_n = j\}$ , a la que suelen llamar *tiempo de retorno* o *tiempo de recurrencia*.

y debe satisfacer

$$\sum_{n=1}^{\infty} f_{ij}^{(n)} \leq 1$$

Existe una importante relación entre las probabilidades de transición y las probabilidades de los tiempos de primera pasada.

Como puede observarse en la definición anterior, cuando  $n = 1$  la probabilidad del tiempo de primera pasada es igual a la probabilidad de transición en un paso.

Para el caso en que  $n > 1$ , el tiempo de primera pasada puede calcularse a partir de las probabilidades de transición de forma recursiva ya que, como sabemos, las probabilidades de transición  $p_{ij}^m$  exploran *todos* los posibles caminos para llegar del estado  $i$  al estado  $j$  en un número determinado de pasos  $m$ .

Por ejemplo,  $p_{ij}^2$  implica dos posibles sucesos: (i) que el tiempo de llegada a  $j$ , partiendo de  $i$ , haya sido igual a 2 o bien (ii) que el tiempo de llegada a  $j$ , partiendo de  $i$ , sea igual a 1 y que luego suceda una transición del estado  $j$  a sí mismo.

Esto es

$$\begin{aligned} p_{ij}^1 &= f_{ij}^{(1)} \\ p_{ij}^2 &= f_{ij}^{(2)} + f_{ij}^{(1)} p_{jj}^1 \\ p_{ij}^3 &= f_{ij}^{(3)} + f_{ij}^{(2)} p_{jj}^1 + f_{ij}^{(1)} p_{jj}^2 \\ &\vdots \end{aligned}$$

Por inducción

$$p_{ij}^n = f_{ij}^{(n)} + \sum_{m=1}^{n-1} f_{ij}^{(m)} p_{jj}^{n-m}, \quad \forall n \geq 2$$

De forma tal que

$$f_{ij}^{(n)} = \begin{cases} p_{ij}^1 & \text{para } n = 1 \\ p_{ij}^n - \sum_{m=1}^{n-1} f_{ij}^{(m)} p_{jj}^{n-m} & \forall n \geq 2 \end{cases} \quad (2.6)$$

Entonces, la probabilidad de que el estado  $i$  alcance *alguna vez* al estado  $j$  se expresa como <sup>4</sup>

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} \quad (2.7)$$

Cuando  $f_{ij} < 1$ , se dice que existe la posibilidad de que el estado  $i$  *nunca* alcance al estado  $j$ . En cambio, cuando  $f_{ij} = 1$  se tiene la certeza de que el estado  $i$  alcanzará *alguna vez* a dicho estado.

Note que cuando  $\sum_{n=1}^{\infty} f_{ij}^{(n)} = 1$  se tiene una distribución de probabilidad para la variable aleatoria  $T_{ij}$ .

Si se desea conocer el tiempo promedio que tarda un estado  $i$  en alcanzar a un estado  $j$  se debe calcular la esperanza de la distribución  $f_{ij}$ .

$\mu_{ij}$  denota el *tiempo esperado de primera pasada* o *tiempo esperado de alcance* del estado  $i$  al estado  $j$  y se define como

$$\mu_{ij} = \begin{cases} \infty & \text{si } \sum_{n=1}^{\infty} f_{ij}^{(n)} < 1 \\ \sum_{n=1}^{\infty} n f_{ij}^{(n)} & \text{si } \sum_{n=1}^{\infty} f_{ij}^{(n)} = 1 \end{cases} \quad (2.8)$$

Cuando  $i = j$  se dice que  $\mu_{ii}$  es el *tiempo esperado de recurrencia*.

Siempre que  $\sum_{n=1}^{\infty} f_{ij}^{(n)} = 1$ ,  $\mu_{ij}$  satisface de manera única

$$\mu_{ij} = 1 + \sum_{k \neq j} p_{ik} \mu_{kj} \quad (2.9)$$

**2.9 Definición.** El *periodo*  $d(i)$  de un estado  $i$  es el número más grande que divide al  $n$  para el cual  $p_{ii}^n > 0$ , i.e.,

$$d(i) = m.c.d. \{n \geq 1 : p_{ii}^n > 0\}$$

**2.10 Definición.** Se dice que el estado  $i$  es *aperiódico* cuando  $d(i) = 1$  y *periódico* si  $d(i) > 1$ <sup>5</sup>.

<sup>4</sup>  $f_{ij} \neq f_{ij}^{(1)}$ .

<sup>5</sup> Si  $p_{ii}^n = 0 \forall n \geq 1$ , decimos que  $d(i) = 0$ .



Con base en las ecuaciones (2.7) y (2.8), y la definición 2.10 los estados de una cadena de Markov se clasifican como sigue:

1. Un estado es *transitorio* si  $f_{ii} < 1$ , i.e., si  $\mu_{ii} = \infty$ .
2. Un estado es *recurrente* si  $f_{ii} = 1$ .
3. Un estado recurrente es *nulo* si  $\mu_{ii} = \infty$  y *no nulo* si  $\mu_{ii} < \infty$ <sup>6</sup>.
4. Un estado recurrente es *ergódico* si es no nulo y aperiódico.

**2.1 Teorema.** *Los estados de una cadena de Markov irreducible son todos del mismo tipo, dado que sólo pueden ser*

- *todos transitorios,*
- *todos recurrentes no nulos, o*
- *todos recurrentes nulos.*

*Incluso, los estados tendrán la misma periodicidad  $n$ .*

**2.4 Ejemplo.** Retomemos el ejemplo del vendedor de calzado de seguridad industrial de la Figura 2.1.

1. Supongamos que ocurrió lo siguiente:

$$\chi_0 = 1, \chi_1 = 3, \chi_2 = 2, \chi_3 = 1, \chi_4 = 2, \chi_5 = 3, \chi_6 = 1, \dots$$

El tiempo de recurrencia:

$$T_{11} = \min\{n \geq 1 : \chi_n = 1 | \chi_0 = 1\} = \min\{3, 6, \dots\} = 3$$

Los tiempos de primera pasada del estado 1 a los demás estados:

$$T_{12} = \min\{n \geq 1 : \chi_n = 2 | \chi_0 = 1\} = \min\{2, 4, \dots\} = 2$$

$$T_{13} = \min\{n \geq 1 : \chi_n = 3 | \chi_0 = 1\} = \min\{1, 5, \dots\} = 1$$

---

<sup>6</sup>Si una cadena de Markov irreducible consta de estados recurrentes *no nulos*, el número de estados es finito.

2. La probabilidad de que el tiempo de primera pasada del estado 3 al estado 1 sea 3:

$$\begin{aligned}
 f_{31}^{(3)} &= p_{31}^3 - \sum_{m=1}^2 f_{31}^{(m)} p_{11}^{3-m} \\
 &= p_{31}^3 - f_{31}^{(1)} p_{11}^2 - f_{31}^{(2)} p_{11}^1 \\
 &= 0.53 - (0.5)(0.56) - (0.35)(0.5) \\
 &= 0.53 - 0.28 - 0.175 \\
 &= 0.075
 \end{aligned}$$

3. Utilizando la ecuación (2.9) se tiene el siguiente sistema de ecuaciones para el tiempo esperado de primera pasada del estado 3 al estado 1:

$$\begin{aligned}
 \mu_{31} &= 1 + p_{32}\mu_{21} + p_{33}\mu_{31} \\
 \mu_{21} &= 1 + p_{22}\mu_{21} + p_{23}\mu_{31} \\
 \mu_{11} &= 1 + p_{12}\mu_{21} + p_{13}\mu_{31}
 \end{aligned}$$

Sustituyendo con los valores de la matriz de transición  $\mathbf{P}$  de la página 40

$$\begin{aligned}
 \mu_{31} &= 1 + 0.5\mu_{21} + 0\mu_{31} \\
 \mu_{21} &= 1 + 0\mu_{21} + 0.3\mu_{31} \\
 \mu_{11} &= 1 + 0.3\mu_{21} + 0.2\mu_{31}
 \end{aligned}$$

Y resolviendo este sistema de ecuaciones tenemos que

$$\mu_{31} = 1.76, \quad \mu_{21} = 1.53$$

y el tiempo esperado de recurrencia del estado 1

$$\mu_{11} = 1.81$$

4. Como  $f_{ij} = 1 \forall i, j \in S$ , la cadena es recurrente.
5. En cuanto a periodicidad se refiere, se tiene que  $p_{11}^n > 0 \forall n$ :

$$p_{11}^1 = 0.5, \quad p_{11}^2 = 0.56, \quad p_{11}^3 = 0.55, \quad p_{11}^4 = 0.5526 \dots$$

Entonces

$$d(1) = m.c.d.\{1, 2, 3, 4, \dots\} = 1$$

y como todos los estados se comunican  $d(1)=d(2)=d(3)$ . Por lo tanto, la cadena es aperiódica.

La cadena es recurrente no nula y aperiódica y, por consecuencia, ergódica.

### 2.1.2. Distribución de estado estable

El caso más interesante para evaluar el comportamiento de una cadena de Markov es aquel en el que la probabilidad  $p_{ij}^m$  no cambia cuando  $m \rightarrow \infty$ .

Dicho de otro modo, cuando existe una probabilidad límite de que la cadena se encuentre en el estado  $j$  después de un número grande de transiciones, sin importar la probabilidad de estado inicial  $\pi_j^{(0)}$ .

**2.11 Definición.** Definimos a la *distribución de probabilidad de estado estable*  $\{\pi_j; j \in S\}$  de una cadena de Markov como

$$\pi_j = \lim_{m \rightarrow \infty} p_{ij}^m$$

Cuando este límite existe, se dice que el estado  $j$  ha alcanzado una *distribución de estado estable*.

La probabilidad de estado estable no implica en forma alguna que la cadena se fije en un estado, ya que ésta sigue teniendo transiciones de un estado a otro con probabilidades  $p_{ij}^n$ .

**2.2 Teorema.** En una cadena de Markov homogénea, irreducible y aperiódica, las probabilidades límite  $\pi_j$  siempre existen y son independientes de la distribución de probabilidad de estado inicial  $\pi_j^{(0)}$ . Aún más, o

1. todos los estados son transitorios o todos los estados son recurrentes nulos. En ambos casos  $\pi_j = 0 \forall j$  y se dice que no hay una distribución de estado estable, o
2. todos los estados son recurrentes no nulos y entonces  $\pi_j > 0 \forall j$ , en cuyo caso el conjunto  $\{\pi_j\}$  es la distribución de probabilidad de estado estable y

$$\pi_j = \frac{1}{\mu_j} \quad (2.10)$$

Las  $\pi_j$  satisfacen de manera única:

$$\sum_j \pi_j = 1 \quad (2.11)$$

$$\pi_j = \sum_i \pi_i p_{ij} \quad (2.12)$$

donde  $\mu_j$  está definida por la ecuación (2.8).

Cuando en una cadena de Markov con espacio de estados  $S$  existe la distribución de probabilidad  $\{\pi_j, j \in S\}$  que satisface la ecuación (2.12), se dice que la cadena es *estacionaria* o que tiene una *distribución estacionaria*.

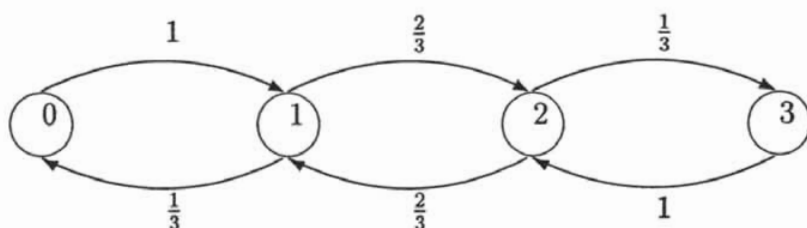
El teorema anterior establece que, bajo ciertas condiciones, la distribución de estado estable es igual a la distribución estacionaria. Sin embargo, esto no se cumple a la inversa, pues puede darse el caso de que una cadena sea estacionaria y que no exista la distribución de estado estable.

La ecuación para  $\pi_j$  puede expresarse de forma matricial. Para ello, definimos al vector de probabilidad  $\Pi$  como

$$\Pi = [\pi_0, \pi_1, \pi_2, \dots]$$

de manera que podamos reescribir la ecuación (2.12) como

$$\Pi = \Pi P \quad (2.13)$$

Figura 2.3: Cadena de Ehrenfest con  $N = 3$ 

La ecuación (2.13) se sigue directamente de la ecuación  $\Pi^{(m)} = \Pi^{(m-1)}\mathbf{P}$  en el límite cuando  $m \rightarrow \infty$ .

El siguiente ejemplo ilustra que no existe una distribución de estado estable para una cadena de Markov periódica.

*2.5 Ejemplo.* La *cadena de Ehrenfest* se originó en el campo de la física como un modelo de dos volúmenes cúbicos de aire conectados a través de un orificio pequeño. En la versión matemática, se tienen dos urnas que contienen un total de  $N$  bolas. Tomamos una de las  $N$  bolas (de cualquiera de las dos urnas) al azar y la depositamos en la otra urna.

Sea  $\chi_n$  el número de bolas que contiene la urna 1 en el  $n$ -ésimo ensayo.

Es lógico que  $\chi_n$  tiene la propiedad markoviana, pues si deseamos saber el número de bolas en la urna 1 al tiempo  $n + 1$ , es claro que la única información relevante de la secuencia observada  $\chi_n, \chi_{n-1}, \chi_{n-2}, \dots, \chi_1, \chi_0$  es  $\chi_n$ , ya que revela el número de bolas en la urna 1 al tiempo antecesor inmediato  $n$ .

Para que el número  $i$  de bolas en la urna 1 se incremente en 1, es necesario tomar una de las  $N - i$  bolas de la urna 2.

$$P[\chi_{n+1} = i + 1 | \chi_n = i] = \frac{N - i}{N}$$

Y evidentemente, el número  $i$  de bolas en la urna 1 se decrementa en 1 con

probabilidad  $\frac{i}{N}$ . Esto es

$$p_{ij} = P[\chi_{n+1} = j | \chi_n = i] = \begin{cases} \frac{N-i}{N} & \text{si } j = i + 1 \\ \frac{i}{N} & \text{si } j = i - 1 \\ 0 & \text{en otro caso} \end{cases}$$

Supongamos  $N = 3$  y una matriz de transición

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 0 & \frac{3}{3} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{3} \\ 0 & 0 & \frac{3}{3} & 0 \end{pmatrix} \end{matrix}$$

Hemos escrito  $\frac{3}{3}$  para enfatizar el patrón que siguen las diagonales de la matriz.

La figura 2.3 muestra las transiciones posibles de un estado a otro.

Para conocer la periodicidad de esta cadena observamos que  $p_{00}^n = 0$  para  $n$  impar, y  $p_{00}^n > 0$  para  $n$  par:

$$p_{00}^2 = \frac{1}{3}, \quad p_{00}^4 = \frac{7}{27}, \quad p_{00}^6 = \frac{61}{243}, \quad p_{00}^8 = \frac{547}{2187} \dots$$

Entonces

$$d(0) = m.c.d.\{2, 4, 6, 8, \dots\} = 2$$

Todos los estados se comunican, por lo tanto  $d(0) = d(1) = d(2) = d(3)$ .

Como  $d(i) > 1$ , la cadena es periódica.

Sea  $\Pi^{(0)} = (1, 0, 0, 0)$ , entonces

$$\begin{aligned} \Pi^{(1)} &= \Pi^{(0)}\mathbf{P} = (0, 1, 0, 0) \\ \Pi^{(2)} &= \Pi^{(1)}\mathbf{P} = (0.33, 0, 0.67, 0) \\ \Pi^{(3)} &= \Pi^{(2)}\mathbf{P} = (0, 0.78, 0, 0.22) \\ \Pi^{(4)} &= \Pi^{(3)}\mathbf{P} = (0.26, 0, 0.74, 0) \\ &\vdots \end{aligned}$$

Es claro que el límite  $\Pi = \lim_{m \rightarrow \infty} \Pi^{(m)}$  no existe.

De manera similar, la cadena debe ser irreducible para que exista una solución única.

*2.6 Ejemplo.* Consideremos ahora el ejemplo de la Figura 2.2, cuya matriz de transición ya hemos planteado en el Ejemplo 2.3. Sea  $\Pi^{(0)} = (1, 0, 0, 0)$ , entonces

$$\begin{aligned}\Pi^{(1)} &= \Pi^{(0)}\mathbf{P} = (0.1, 0.4, 0.5, 0) \\ \Pi^{(2)} &= \Pi^{(1)}\mathbf{P} = (0.11, 0.04, 0.7, 0.15) \\ \Pi^{(3)} &= \Pi^{(2)}\mathbf{P} = (0.151, 0.044, 0.445, 0.36) \\ \Pi^{(4)} &= \Pi^{(3)}\mathbf{P} = (0.1041, 0.0604, 0.342, 0.4935) \\ &\vdots\end{aligned}$$

Si se continúa calculando  $\Pi^{(m)}$ , se observa que el límite  $\Pi = \lim_{m \rightarrow \infty} \Pi^{(m)}$  no existe.

Ahora, calculemos la distribución de estado estable para el caso de nuestro amigo Homobono.

*2.7 Ejemplo.* Dadas las características de la cadena de la Figura 2.1, la distribución de estado estable se calcula a partir de la distribución estacionaria. Usando la ecuación (2.13) podemos escribir las siguientes ecuaciones lineales:

$$\begin{aligned}\pi_1 &= 0.5\pi_1 + 0.7\pi_2 + 0.5\pi_3 \\ \pi_2 &= 0.3\pi_1 + 0\pi_2 + 0.5\pi_3 \\ \pi_3 &= 0.2\pi_1 + 0.3\pi_2 + 0.5\pi_3\end{aligned}\tag{2.14}$$

Note que en el sistema de ecuaciones (2.14), la primera ecuación es una combinación lineal de la segunda y la tercera, lo cual indica que existe una dependencia lineal entre ellas.

Siempre habrá una dependencia lineal entre las ecuaciones del sistema que se desprende de la ecuación (2.13) ya que se tendrá un sistema de  $N + 2$  ecuaciones con  $N + 1$  incógnitas y solución única, por lo cual una de las ecuaciones será redundante y podrá eliminarse del sistema, y en su lugar utilizaremos la ecuación (2.11) para encontrar la solución.

Haciendo uso de la ecuación (2.11), de dos ecuaciones cualesquiera del sistema

(2.14) y resolviendo este nuevo sistema tenemos que:

$$\pi_1 = 0.552$$

$$\pi_2 = 0.260$$

$$\pi_3 = 0.188$$

De manera que es más probable encontrar a Homobono en el Distrito Federal. De hecho, él regresa cada 1.81 semanas al D.F., pues despejando  $\mu_j$  de la ecuación (2.10) tenemos que

$$\mu_1 = \frac{1}{\pi_1} = \frac{1}{0.552} = 1.81$$

lo cual coincide con el resultado obtenido en la página 49.

### 2.1.3. Cadenas absorbentes y comportamiento transitorio

Cuando hablamos de cadenas de Markov para modelar sistemas reales, frecuentemente es útil conocer el número de pasos (o, equivalentemente, el tiempo) que la cadena permanecerá en estados transitorios antes de alcanzar a un estado absorbente.

Sea una cadena de Markov que consiste en un conjunto  $S_t$  de  $n_t$  estados transitorios y en un conjunto  $S_a$  de  $n_a$  estados absorbentes, sin perder de vista que una clase recurrente puede ser considerada en su conjunto como un estado absorbente.

Comenzaremos nuestro análisis numerando los estados de la cadena de Markov de manera que los estados absorbentes sean los primeros elementos del espacio muestral  $S$  y escribiendo la matriz de transición  $\mathbf{P}$  como

$$\mathbf{P} = \begin{pmatrix} I & 0 \\ R & Q \end{pmatrix} \quad (2.15)$$

Una vez que la cadena se encuentra en un estado absorbente nunca saldrá de ahí, por lo que  $I$  es la matriz identidad con todos los elementos  $p_{ii} = 1$ ,  $1 \leq i \leq n_a$ .  $R$  es una matriz de  $n_t \times n_a$  que describe el movimiento de la



cadena de los estados transitorios a los estados absorbentes; y  $Q$ , una matriz de  $n_t \times n_t$  que describe el movimiento de la cadena en los estados transitorios. Como no es posible pasar de un estado absorbente a uno transitorio 0 es la matriz de ceros de tamaño  $n_a \times n_t$ .

Dado que la fórmula para la multiplicación de matrices prevalece para matrices escritas por bloques, podemos calcular las potencias de  $\mathbf{P}$  en términos de las matrices  $R$  y  $Q$ :

$$\mathbf{P}^2 = \begin{pmatrix} I & 0 \\ R + QR & Q^2 \end{pmatrix}$$

y

$$\mathbf{P}^3 = \begin{pmatrix} I & 0 \\ R + QR + Q^2R & Q^3 \end{pmatrix}$$

o en general,

$$\mathbf{P}^n = \begin{pmatrix} I & 0 \\ N_n R & Q^n \end{pmatrix}$$

donde  $N_n = I + Q + Q^2 + \dots + Q^{n-1} = \sum_{i=1}^n Q^{i-1}$ .

**2.3 Teorema.** Cuando  $n \rightarrow \infty$ , entonces  $Q^n \rightarrow 0$  y  $N_n \rightarrow (I - Q)^{-1}$ . En particular, la matriz  $I - Q$  es invertible.

La primera parte del teorema anterior es intuitiva, pues sabemos que para un estado transitorio  $\pi_j = 0$ . La demostración puede consultarse en Çinlar 1997 [9].

Se dice que la matriz

$$N = [n_{ij}] = (I - Q)^{-1}$$

es la *matriz fundamental* de la cadena de Markov.

Del último teorema se sigue que

$$\lim_{m \rightarrow \infty} \mathbf{P}^m = \begin{pmatrix} I & 0 \\ NR & 0 \end{pmatrix} \quad (2.16)$$

En el caso de cadenas absorbentes, los únicos estados iniciales interesantes son los transitorios.

La variable aleatoria  $v_{ij}$  denota el número de visitas del estado  $i$  al estado  $j$ . Si  $i, j \in S_a$  o si  $i \in S_t$  y  $j \in S_a$ ,  $E[v_{ij}] = \infty$ . En cambio, si  $i \in S_a$  y  $j \in S_t$  o si  $i$  y  $j$  son absorbentes, pero pertenecen a clases distintas,  $E[v_{ij}] = 0$ .

Supongamos ahora que comenzamos en un estado inicial  $i \in S_t$ . Para cada estado  $j \in S_t$ ,  $v_{ij}$  denota el número de visitas al estado  $j$  antes de que un estado absorbente sea alcanzado. Cuando  $i = j$ ,  $v_{ij} = 1$ .

Por el Teorema 2.2 sabemos que  $v_{ij} < \infty$  para todo estado transitorio  $j$ , y que  $v_{ij}$  tiene por tanto un valor esperado finito.

**2.4 Teorema.** Para todo par de estados transitorios  $i, j$

$$E[v_{ij}] = n_{ij}$$

donde la matriz  $N$  es la matriz fundamental.

*Demostración.* Por la propiedad markoviana tenemos que

$$E[v_{ij}] = \delta_{ij} + \sum_{k \in S_t} q_{ik} E[v_{kj}]$$

donde  $\delta_{ij}$  es la función *delta de Kronecker*:

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{en otro caso} \end{cases}$$

y  $q_{ik}$  es la probabilidad de transición del estado *transitorio*  $i$  al estado *transitorio*  $k$ . Esta ecuación reconoce el hecho de que la *primera transición*, desde el estado  $i$ , puede ser al estado  $j$  o a algún otro estado  $k$ .

La primera parte de la ecuación establece que si la primera transición ocurre del estado  $i$  al estado  $j$ , el *número de visitas* antes de alcanzar un estado absorbente es 1.

La segunda parte contempla *todas* las posibilidades de que la primera transición ocurra del estado  $i$  al estado  $k$  ( $k \neq j$ ), lo que ocurre con probabilidad  $q_{ik}$ , y

el número esperado de visitas del estado  $k$  al estado  $j$  antes de alcanzar un estado absorbente.

Denotemos como  $M$  la matriz cuyos elementos sean  $E[v_{ij}]$  para toda  $i, j \in S_t$ . Entonces, la última ecuación puede reexpresarse como

$$M = I + QM$$

de manera que  $M = (I - Q)^{-1} = N$ . □

Sea  $\mathbf{V} = [E[v_{ij}]]$ , es decir, la matriz cuyos elementos son el número esperado de visitas de un estado a otro de la cadena.

El número *total* de pasos antes de alcanzar un estado absorbente es la suma de todas las veces que visitamos a cada estado en  $S_t$  antes de la absorción. Sea  $v_i$  la variable aleatoria que denota este número total de pasos y  $\tau_i$  su valor esperado.

### 2.5 Teorema.

$$\tau_i = E[v_i] = \sum_{j \in S_t} n_{ij} \quad i \in S_t \quad (2.17)$$

y  $\tau_i < \infty$ .

*Demostración.* Dado que la esperanza de la suma es la suma de las esperanzas, este resultado se sigue del teorema anterior. □

*2.8 Ejemplo.* Supongamos que tenemos la siguiente matriz de transición.

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{array}{l} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} \left( \begin{array}{ccccccc} 0.2 & 0.8 & 0 & 0 & 0 & 0 & 0 \\ 0.7 & 0.3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.3 & 0.5 & 0.2 & 0 & 0 \\ 0 & 0 & 0.6 & 0 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0.4 & 0.6 & 0 & 0 \\ 0 & 0.1 & 0.1 & 0.2 & 0.2 & 0.3 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0 & 0.1 & 0.2 & 0.4 \end{array} \right) \end{array}$$

Después de hacer un análisis detallado podemos observar que en esta cadena de Markov las clases  $C[1] = \{1, 2\}$  y  $C[3] = \{3, 4, 5\}$  son recurrentes, mientras que la clase  $C[6] = \{6, 7\}$  es transitoria.

Haciendo  $a = C[1]$ ,  $b = C[3]$ ,  $c = 6$  y  $d = 7$ , podemos escribir la matriz de transición de acuerdo a la ecuación (2.15):

$$P = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{cccc} & a & b & c & d \\ \left( \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0.1 & 0.5 & 0.3 & 0.1 \\ 0.2 & 0.2 & 0.2 & 0.4 \end{array} \right) \end{array}$$

De manera que

$$R = \begin{pmatrix} 0.1 & 0.5 \\ 0.2 & 0.2 \end{pmatrix}, \quad Q = \begin{pmatrix} 0.3 & 0.1 \\ 0.2 & 0.4 \end{pmatrix}$$

y

$$N = (I - Q)^{-1} = \begin{pmatrix} 0.7 & -0.1 \\ -0.2 & 0.6 \end{pmatrix}^{-1} = \begin{pmatrix} 1.5 & 0.25 \\ 0.5 & 1.75 \end{pmatrix}$$

Entonces

$$V = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{cccc} & a & b & c & d \\ \left( \begin{array}{cccc} \infty & 0 & 0 & 0 \\ 0 & \infty & 0 & 0 \\ \infty & \infty & 1.5 & 0.25 \\ \infty & \infty & 0.5 & 1.75 \end{array} \right) \end{array}$$

y por la ecuación (2.17)

$$\tau_c = 1.5 + 0.25 = 1.75$$

$$\tau_d = 0.5 + 1.75 = 2.25$$

$$NR = \begin{pmatrix} 1.5 & 0.25 \\ 0.5 & 1.75 \end{pmatrix} \begin{pmatrix} 0.1 & 0.5 \\ 0.2 & 0.2 \end{pmatrix} = \begin{pmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{pmatrix}$$

Y por último

$$\lim_{m \rightarrow \infty} \mathbf{P}^m = \begin{array}{c} a \\ b \\ c \\ d \end{array} \begin{array}{cccc} & a & b & c & d \\ \left( \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \end{array} \right) \end{array}$$

Lo cual quiere decir que las probabilidades de que en el largo plazo, las clases  $a$  y  $b$  absorban al estado  $c$  son 0.2 y 0.8, respectivamente.

De igual forma, las probabilidades de que en el largo plazo, las clases  $a$  y  $b$  absorban al estado  $d$  son 0.4 y 0.6, respectivamente.

Para conocer las probabilidades de absorción y la distribución de estado estable para *cada uno* de los estados que conforman a la cadena del ejemplo anterior, se recomienda al lector revisar el Capítulo 6 de Çinlar 1997 [9], en el que se trata el tema de forma exhaustiva. En dicha obra también se hace un análisis riguroso para el caso de cadenas markovianas periódicas.

Con esto daremos por terminado nuestro estudio sobre cadenas de Markov.

## 2.2. Procesos de Markov

Los *procesos de Markov* son cadenas de Markov con espacio parametral de tiempo continuo. En 1923 Norbert Wiener fue el primero en darle un tratamiento riguroso al caso continuo de las cadenas markovianas. La teoría general de dichos procesos fue fundamentada finalmente por Andrei Kolmogorov en 1930.

En la sección anterior Homobono debía decidir cada domingo si quedarse o abandonar el lugar en donde se encontraba. Esto es, debía tomar una decisión en intervalos de tiempo regulares e iguales.

Supongamos que ahora le permitimos a nuestro amigo decidir en un *tiempo arbitrario* a dónde ir, es decir, ahora puede tomar dicha decisión a cualquier hora

de cualquier día de la semana. Tenemos pues la versión con espacio parametral de tiempo continuo del ejemplo anterior.

**2.12 Definición.** El proceso estocástico  $\{\chi_t, t \geq 0\}$  es una cadena de Markov de tiempo continuo o bien, un *proceso de Markov* si

$$P[\chi_{t+s} = j | \chi_s = i, \chi_{s_1} = i_1, \dots, \chi_{s_0} = i_0] = P[\chi_t = j | \chi_0 = i] \quad (2.18)$$

para cualquier secuencia  $s_0, s_1, \dots, s_n$  tal que  $0 \leq s_0 < s_1 < \dots < s_n$  y los posibles estados  $i_0, \dots, i_n, i, j \in S$  donde  $S$  es el espacio de estados discreto finito o infinito numerable.

**2.13 Definición.** Un proceso de Markov es *homogéneo* si

$$P[\chi_{t+s} = j | \chi_s = i] = P[\chi_t = j | \chi_0 = i] = p_{ij}^t \quad \text{con } s \neq 0$$

es decir, si  $P[\chi_{t+s} = j | \chi_s = i]$  es independiente de  $s$ .

Para nuestro estudio consideraremos sólo procesos markovianos homogéneos.

El segundo miembro de la ecuación ( 2.18) es la probabilidad de transición del proceso de Markov y la escribimos

$$p_{ij}^t = P[\chi_t = j | \chi_0 = i]$$

que denota la probabilidad de que el proceso estará en el estado  $j$  al tiempo  $t$ , dado que está en el estado  $i$  al tiempo 0.

Note que necesitamos definir

$$p_{ij}^t = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{en otro caso} \end{cases}$$

para establecer el hecho de que el proceso no abandonará inmediatamente al estado en cuestión para llegar a otro.

Ahora supongamos que un proceso de Markov llega al estado  $i$  al tiempo 0, y que el proceso no deja al estado  $i$  durante las siguientes  $s$  unidades de tiempo. Del hecho de que el proceso se encuentre en el estado  $i$  al tiempo  $s$  se sigue, por la propiedad markoviana, que la probabilidad de que se quede en dicho estado

en el intervalo de tiempo  $[s, s + t]$ , es la probabilidad de que permanezca en el estado  $i$  por al menos  $t$  unidades de tiempo. Esto es, si  $\gamma_i$  denota el tiempo que el proceso permanece en el estado  $i$  antes de hacer una transición a cualquier otro estado, entonces

$$P[\gamma_i > t + s | \gamma_i > s] = P[\gamma_i > t] \quad \text{para toda } s, t \geq 0$$

Esto significa que la variable aleatoria  $\gamma_i$  no tiene memoria y la única distribución de probabilidad continua que satisface dicha condición es la exponencial. Esta distribución tiene un sólo parámetro, llámese  $\lambda$ , donde la media es  $\frac{1}{\lambda}$  y la función de densidad de probabilidad es

$$f_{\gamma_i}(t) = \lambda e^{-\lambda t}, \quad \text{para } t \geq 0$$

y calculando su función de distribución acumulativa

$$F_{\gamma_i}(t) = P[\gamma_i \leq t] = 1 - e^{-\lambda t}, \quad \text{para } t \geq 0$$

obtendremos la probabilidad de que el proceso no deje al estado  $i$  durante las siguientes  $t$  unidades de tiempo.

Este resultado lleva a una forma equivalente para describir un proceso markoviano:

1. La variable aleatoria  $\gamma_i$  tiene una distribución exponencial con media  $\frac{1}{\lambda_i}$ .
2. Cuando el proceso sale de un estado  $i$ , el proceso se mueve a otro estado  $j$ , con probabilidad  $p_{ij}$ , en donde

$$p_{ij}^t \geq 0, \quad i, j = 1, 2, \dots, N; t \geq 0$$

y

$$\sum_{j \in S} p_{ij}^t = 1, \quad i = 1, 2, \dots, N; t \geq 0$$

3. El siguiente estado que se visita después del estado  $i$  es independiente del tiempo que pasó en el estado  $i$ .

El objetivo de los procesos markovianos es encontrar la probabilidad de que el proceso esté en el estado  $i$  al tiempo  $t$  o bien conocer la distribución de estado

estable. En el caso de los procesos markovianos, la solución es distinta a la del caso discreto indicada en el teorema ( 2.2), aun cuando los argumentos son análogos.

Igual que las probabilidades de transición en un paso jugaron un papel primordial al describir una cadena de Markov, el papel análogo para los procesos de Markov lo tienen las tasas de transición.

**2.14 Definición.** Si el límite

$$q_{ij} = \lim_{h \rightarrow 0} \frac{p_{ij}^h}{h} \quad \text{para } j \neq i$$

existe,  $q_{ij}$  es la *intensidad* o *tasa de transición* de  $i$  a  $j$ .

Utilizando los mismos argumentos que en la página 37, de que el proceso visita a un estado intermedio  $k$  al tiempo  $h$  en el recorrido que hace para pasar del estado  $i$  al estado  $j$  obtenemos la ecuación de *Chapman-Kolmogorov* para los procesos markovianos:

$$p_{ij}^{t+h} = \sum_{k \in S} p_{ik}^h p_{kj}^t \quad (2.19)$$

Restando a la ecuación de Chapman-Kolmogorov el elemento  $p_{ij}^t$  tenemos

$$\begin{aligned} p_{ij}^{t+h} - p_{ij}^t &= \left( \sum_{k \in S} p_{ik}^h p_{kj}^t \right) - p_{ij}^t \\ &= \left( \sum_{\{k \neq i\} \in S} p_{ik}^h p_{kj}^t \right) + [p_{ii}^h - 1] p_{ij}^t \quad (2.20) \end{aligned}$$

Con esto nuestro objetivo es dividir cada miembro de la ecuación anterior por  $h$  y hacer que  $h \rightarrow 0$  para poder calcular

$$p_{ij}^{t'} = \lim_{h \rightarrow 0} \frac{p_{ij}^{t+h} - p_{ij}^t}{h}$$



Ignorando el detalle de intercambiar el límite y la suma y por la definición de tasa de transición se tiene que

$$\lim_{h \rightarrow 0} \frac{1}{h} \sum_{\{k \neq i\} \in S} p_{ik}^h p_{kj}^t = \sum_{\{k \neq i\} \in S} q_{ik} p_{kj}^t$$

Para el otro término, hacemos notar que  $1 - p_{ii}^h = \sum_{\{k \neq i\} \in S} p_{ik}^h$  de modo que

$$\lim_{h \rightarrow 0} \frac{p_{ii}^h - 1}{h} = - \lim_{h \rightarrow 0} \sum_{\{k \neq i\} \in S} \frac{p_{ik}^h}{h} = - \sum_{\{k \neq i\} \in S} q_{ik} = -\lambda_i$$

y

$$\lim_{h \rightarrow 0} \frac{p_{ii}^h - 1}{h} p_{ij}^t = -\lambda_i p_{ij}^t$$

Combinando todos estos hechos y partiendo de la ecuación ( 2.20) y de la definición de derivada tenemos

$$p_{ij}^{t'} = \sum_{\{k \neq i\} \in S} q_{ik} p_{kj}^t - \lambda_i p_{ij}^t \quad (2.21)$$

Para simplificar la última expresión introducimos a la matriz  $\mathbf{Q}$ , conocida como la *matriz de tasas de transición*<sup>7</sup>

$$\mathbf{Q} = \begin{cases} q_{ij} & \text{si } j \neq i \\ -\lambda_i & \text{si } j = i \end{cases}$$

y las matrices  $\mathbf{P}^t = [p_{ij}^t]$  y  $\mathbf{P}^{t'} = [p_{ij}^{t'}]$

Note que los elementos  $q_{ij}$  de la matriz  $\mathbf{Q}$ , es decir, los elementos que no pertenecen a la diagonal de esta matriz son no negativos, mientras que los elementos de la diagonal  $\lambda_i$  son negativos. De este modo, la suma de cada renglón de la matriz de transición  $\mathbf{Q}$  es igual a cero.

De manera que usando notación matricial podemos reescribir ( 2.21) simplemente como

$$\mathbf{P}^{t'} = \mathbf{Q}\mathbf{P}^t \quad (2.22)$$

<sup>7</sup>Que no debe de confundirse con la matriz  $Q$  de  $n_t \times n_t$  mencionada en la subsección 2.1.3, la cual describe el movimiento de una cadena en estados transitorios.

A esta ecuación se le conoce como la *ecuación diferencial de Chapman-Kolmogorov hacia atrás*<sup>8</sup> debido a que para calcular las probabilidades de transición del estado  $i$  al estado  $j$  al tiempo  $t+h$ , condicionamos sobre el estado intermedio  $k$  al tiempo  $h$ .

La interpretación intuitiva de  $\lambda_i$  y  $q_{ij}$  es la siguiente. En particular,  $\lambda_i$  es la *tasa de transición hacia afuera del estado  $i$*  en el sentido de que  $\lambda_i$  es el número esperado de veces que el proceso deja al estado  $i$  por unidad de tiempo que pasa en el estado  $i$ . De manera similar,  $q_{ij}$  es *tasa de transición del estado  $i$  al estado  $j$*  en el sentido de que  $q_{ij}$  es el número esperado de veces que el proceso transita del estado  $i$  al estado  $j$  por unidad de tiempo que pasa en el estado  $i$ .

Si hacemos el mismo análisis, pero ahora condicionando sobre el estado intermedio  $k$  al tiempo  $t$

$$\begin{aligned} p_{ij}^{t+h} - p_{ij}^t &= \left( \sum_{k \in S} p_{ik}^t p_{kj}^h \right) - p_{ij}^t \\ &= \left( \sum_{\{k \neq i\} \in S} p_{ik}^t p_{kj}^h \right) + [p_{ii}^h - 1] p_{ij}^t \end{aligned}$$

y haciendo los mismos cálculos que acabamos de señalar obtenemos que

$$p_{ij}^{t'} = \sum_{\{k \neq i\} \in S} p_{ik}^t q_{kj} - p_{ij}^t \lambda_j \quad (2.23)$$

Nuevamente, introduciendo la notación matricial, podemos escribir

$$P^{t'} = P^t Q \quad (2.24)$$

Y de esta forma hemos obtenido la *ecuación diferencial de Chapman-Kolmogorov hacia adelante*.

Comparando (2.24) con (2.22) podemos apreciar que  $P^t Q = Q P^t$  y que estas dos formas de las ecuaciones diferenciales de Chapman-Kolmogorov sólo se distinguen la una de la otra en la escritura de la matriz de tasas de transición  $Q$  a la izquierda o a la derecha. Recordemos que, en general, la multiplicación

<sup>8</sup>En algunos textos se omite el nombre de Chapman para esta ecuación diferencial.

de matrices no satisface la conmutatividad, es decir,  $\mathbf{AB} \neq \mathbf{BA}$ . Por este motivo enfatizamos que  $\mathbf{P}^t\mathbf{Q} = \mathbf{QP}^t$ .

Las ecuaciones diferenciales de Chapman-Kolmogorov permiten el cálculo de las probabilidades de transición  $p_{ij}^t$  a partir de las tasas de transición  $q_{ij}$ . Sin embargo, cabe señalar que, en general, el cálculo de las  $p_{ij}^t$  aun por medio de las ecuaciones de Chapman-Kolmogorov es una tarea complicada.

Antes de concluir esta sección es necesario señalar que la clasificación de estados para los procesos markovianos es exactamente la misma que mencionamos en la subsección 2.1.1 para las cadenas de Markov.

### 2.2.1. Distribución de estado estable

El estudio del comportamiento límite de los procesos markovianos es más sencillo que en el caso de las cadenas markovianas debido a que no es necesario preocuparse por la periodicidad.

**2.6 Teorema.** *Un proceso markoviano  $\chi_t$  irreducible tiene distribución de estado estable si*

$$\pi_j = \lim_{t \rightarrow \infty} p_{ij}^t$$

Al establecer que  $\chi_t$  es irreducible nos referimos a que para dos estados cualesquiera del proceso  $x$  y  $y$  es posible llegar de  $x$  a  $y$  en un número finito de transiciones. Para ser precisos, que hay una secuencia de estados  $x_0 = x, x_1, \dots, x_n = y$  tales que  $q_{x_{m-1}x_m} > 0$  para  $1 \leq m \leq n$ .

Cuando el proceso markoviano es irreducible, homogéneo y todos los estados que lo conforman son recurrentes no nulos este límite existe y es independiente de la distribución de estado inicial. La distribución límite en este caso se determina de forma análoga al teorema 2.2.

En el caso de las cadenas markovianas la distribución estacionaria es la solución de  $\Pi\mathbf{P} = \Pi$ . Como no hay una primera  $t > 0$ , en los procesos markovianos se dice que  $\Pi$  es el vector de probabilidades estacionarias si  $\Pi\mathbf{P}^t = \Pi$  para toda  $t > 0$ , lo cual resulta sumamente difícil de verificar considerando que esta

expresión involucra a todas las  $P^t$  y la complejidad que implica el cálculo de cada matriz  $P^t$ .

El siguiente teorema resuelve este problema al expresar a  $\Pi$  en términos de la matriz de tasas de transición.

**2.7 Teorema.**  $\Pi$  es un vector de probabilidades estacionarias si, y sólo si,  $\Pi Q = 0$ .

*Demostración.* La ecuación ( 2.24) señala que  $P^{t'} = P^t Q$ . Llevando esta expresión matricial a su forma algebraica, multiplicando por  $\pi_i$  y sumando

$$\sum_{i \in S} \pi_i p_{ij}^{t'} = \sum_{i, k \in S} \pi_i p_{ik}^t Q_{kj}$$

Extrayendo la derivada de la suma, y usando  $\Pi P' = \Pi$ , el primer miembro de esta ecuación puede escribirse

$$\frac{d}{dt} \left( \sum_{i \in S} \pi_i p_{ij}^t \right) = \frac{d}{dt} \pi_j = 0$$

mientras que al sumar sobre  $i$  en el segundo miembro de la ecuación tenemos que

$$\sum_{i, k \in S} \pi_i p_{ik}^t = \pi_k$$

De estos dos últimos hechos se depende que

$$0 = \sum_{k \in S} \pi_k Q_{kj} = \pi_j Q_j$$

donde  $\pi_j Q_j$  es el  $j$ -ésimo componente del vector  $\Pi Q$ .

Esto quiere decir que  $\pi_j Q_j$  es la tasa a la que el proceso *deja* al estado  $j$ , ya que  $\pi_j$  es la probabilidad de que el proceso esté en el estado  $j$  y  $Q_j$  es la tasa de transición hacia afuera de  $j$  dado que el proceso se encuentra en el estado  $j$ . De manera similar,  $\pi_k Q_{kj}$  es la tasa a la que el proceso *entra* al estado  $j$  desde el estado  $k$ , ya que  $Q_{kj}$  es la tasa de transición del estado  $k$  al estado  $j$  dado que el proceso se encuentra en el estado  $k$ . Sumando sobre toda  $k \neq j$ ,  $\pi_k Q_{kj}$  proporciona la tasa a la que el proceso entra al estado  $j$  desde cualquier

estado. Por esta razón, la ecuación anterior establece que la tasa a la cual el proceso deja al estado  $j$  debe ser igual a la tasa a la que el proceso entra en el estado  $j$ .

Hasta este momento hemos mostrado que si  $\Pi$  es estacionario, entonces  $\Pi Q = 0$ . Si ahora comenzamos intercambiando la derivada y la suma podemos concluir que

$$\frac{d}{dt} \left( \sum_{i \in S} \pi_i p_{ij}^t \right) = \sum_{i \in S} \pi_i p_{ij}^{t \prime}$$

Para evaluar el segundo miembro de esta ecuación utilizaremos ahora la ecuación diferencial de Chapman-Kolmogorov (2.21). Multiplicándola por  $\pi_i$  y sumando se tiene que

$$\frac{d}{dt} \sum_{i \in S} \pi_i p_{ij}^t \prime = \sum_{k \in S} \pi_k Q_k p_{kj}^t = 0$$

lo que se sigue debido a que hemos asumido de  $\Pi Q = 0$ . Como la derivada es cero, los elementos de  $\Pi P^t$  son constantes y deben ser entonces iguales a los elementos de  $\Pi$ .  $\square$

Por supuesto, las  $\pi_j$  satisfacen

$$\sum_j \pi_j = 1$$

Con esto concluimos nuestra introducción a los procesos markovianos. Para ahondar en la teoría markoviana se sugieren los libros de Ross 1996 [30] y Parzen 1962 [27] los cuales tratan el tema con un nivel teórico más elevado, y Durrett 1999 [13], este último con un enfoque más práctico y una serie de ejemplos que ilustran las definiciones expuestas.

## Capítulo 3

# Redes de Petri

*“Un Matemático es un Quijote moderno que lucha en un mundo real con armas imaginarias”.*

P. Corcho

Las redes de Petri fueron creadas en 1962 por Carl Adam Petri para su Tesis doctoral titulada *Kommunikation mit Automaten* (Comunicación con autómatas) en la cual establece los fundamentos teóricos de las mismas.

La importancia de las redes de Petri radica en dos aspectos:

- (i) Permiten la descripción de la concurrencia y sincronización inherente de sistemas modernos.
- (ii) Pueden ser analizadas de manera formal y brindar información del comportamiento dinámico del sistema modelado, es decir, podemos modelar no sólo la estructura de un sistema, sino además su comportamiento, y llevar el modelo a condiciones límite, que en un sistema real son difíciles de lograr o muy costosas.

Para comprender la razón por la cual las redes de Petri son una herramienta útil para el estudio de sistemas que involucran concurrencia y sincronización, es necesario conocer la estructura y funcionamiento de un sistema.

Un *sistema* se compone de módulos que interactúan entre sí, los cuales pueden ser considerados por sí mismos un sistema. Podemos analizar a cada módulo por separado, pero sin perder de vista la interacción que guarda con los demás módulos. Ahora bien, un sistema tiene variaciones en el transcurso del tiempo y no permanece estático. Las variaciones que sufre un sistema están determinadas por las variaciones que sufren a su vez cada uno de los módulos, las cuales se deben a eventos que se presentan en los mismos. Para que un evento ocurra es necesario que se cumplan ciertas condiciones (condiciones previas) y la ocurrencia de dicho evento nos conduce a nuevas condiciones (condiciones posteriores).

Por último, se dice que un sistema es concurrente si múltiples eventos *independientes* ocurren en paralelo, es decir, al mismo tiempo. Al ser esto posible, y considerando la interacción entre módulos, es necesaria la sincronización de los eventos.

Para modelar un sistema a través de las redes de Petri debemos primeramente identificar no sólo sus condiciones y eventos, sino además la interacción de sus módulos, de manera tal que podamos hacer la analogía entre el sistema y el modelo.

En este capítulo analizaremos la estructura, comportamiento y propiedades de las redes de Petri ordinarias.

### 3.1. Representación gráfica

Además de modelar y analizar formalmente sistemas que requieran de concurrencia y sincronización, las redes de Petri ofrecen una representación gráfica del sistema a modelar. En general, la representación gráfica de las redes de Petri está conformada por los siguientes elementos:

**Lugares**, representados por círculos. Estos modelan condiciones u objetos.

**Elementos**, dibujados como puntos negros. Representan el valor específico de la condición u objeto.

**Transiciones**, representadas por rectángulos. Estas modelan las actividades que producen los cambios en los valores de las condiciones y objetos.

**Arcos**, que especifican la interconexión de lugares y transiciones indicando además qué objetos cambian por una determinada actividad.

Las redes de Petri son *gráficas bipartitas*, es decir, podemos conectar un lugar a una transición y viceversa, pero nunca dos lugares (o dos transiciones) entre sí.

Generalmente, al utilizar las redes de Petri para modelar un sistema, los lugares representan condiciones y las transiciones representan eventos.

## 3.2. Estructura

Las redes de Petri ordinarias, también conocidas como redes *Lugar-Transición*, son gráficas bipartitas dirigidas compuestas por lugares, representados por círculos, y transiciones, representadas por rectángulos.

**3.1 Definición.** Sea  $V$  un conjunto finito no vacío y sea  $A \subseteq V \times V$ . Al par  $G = (V, A)$  se le llama *gráfica dirigida* (en  $V$ ), donde  $V$  es el conjunto de *vértices*, o *nodos*, y  $A$  es el conjunto de *arcos*.

Usualmente, las gráficas dirigidas son determinadas por la descripción de sus elementos (en el caso de las redes de Petri ordinarias existen dos clases de nodos), y funciones o matrices que especifican su interconexión. Por lo pronto utilizaremos la notación correspondiente a funciones y de aquí en adelante nos referiremos a las redes de Petri ordinarias simplemente como redes de Petri.

**3.2 Definición.** *Red de Petri* ( $PN$ ) es la 4-tupla  $PN = (P, T, I^-, I^+)$  donde

- $P = \{p_1, \dots, p_n\}$  es un conjunto de lugares finito y no vacío.
- $T = \{t_1, \dots, t_m\}$  es un conjunto de transiciones finito y no vacío.
- $P \cap T = \emptyset$ .



- $I^-, I^+ : P \times T \rightarrow \mathbb{N}$  son las funciones de incidencia hacia atrás y hacia adelante, respectivamente.

**3.3 Definición.** *Red de Petri marcada* es la 5-tupla  $PN = (P, T, I^-, I^+, M_0)$  donde  $M_0 : P \rightarrow \mathbb{N}$  es el marcado inicial.

Considerando únicamente la estructura de las redes de Petri, éstas pueden ser consideradas como una 4-tupla  $PN = (P, T, I^-, I^+)$  omitiendo el marcado inicial  $M_0$ . El marcado  $M_0$  representa el conjunto de condiciones iniciales de un sistema. Algunos autores llaman a las redes no marcadas  $PN = (P, T, I^-, I^+)$  redes de Petri; y a las redes de Petri marcadas  $PN = (P, T, I^-, I^+, M_0)$ , sistemas.

Las funciones  $I^-$  e  $I^+$  especifican la conexión entre lugares y transiciones. Si  $I^-(p, t) > 0$ , quiere decir que un arco lleva del lugar  $p$  a la transición  $t$ , por lo cual a  $I^-$  se le llama función de incidencia hacia atrás de la transición  $t$ .  $I^-$  hace un mapeo a los números naturales atendiendo al peso del arco que lleva de  $p$  a  $t$ .

En la representación gráfica de las redes de Petri el peso del arco se escribe cerca del correspondiente arco. Por convención, cuando no se especifica cuál es el peso del arco se entenderá que éste es igual a 1. Los pesos asignados a los arcos especifican que la transición estará activada sólo cuando haya, al menos, el número de elementos señalados por el peso del arco en el lugar  $p$  que lleva a dicha transición. Al activarse la transición se destruirá exactamente el mismo número de elementos de  $p$ . De manera similar,  $I^+(p, t)$  especifica el número de elementos creados en el lugar  $p$  en caso de que se active  $t$ .

**3.1 Ejemplo.** Después de años de ardua labor, Homobono Buenrostro decidió iniciar su propio negocio de venta de equipo de seguridad industrial. En su negocio Homobono cuenta con un sistema que monitorea y autoriza el acceso a la base de datos en donde se resume la información más relevante. Al acceder a la base de datos sólo es posible consultar datos o capturar nuevos datos, pues nuestro amigo ha tomado la precaución de asegurarse de que sólo él pueda modificar la estructura de la base de datos. En el negocio Homobono cuenta con tres ayudantes y una computadora. Es obvio que varios ayudantes pueden estar leyendo al mismo tiempo la información resumida en la base de datos, pero sólo uno puede capturar información a la vez.

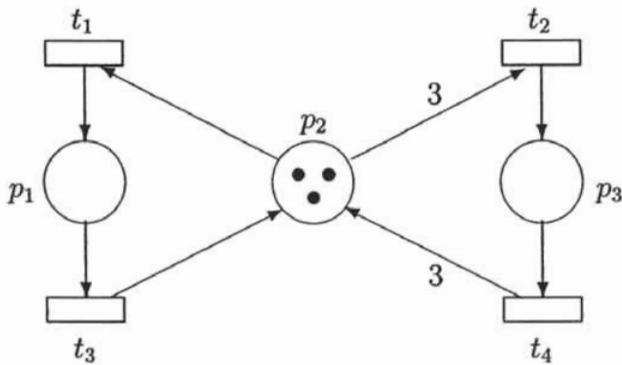


Figura 3.1: Red de Petri

Primeramente, identifiquemos los eventos y condiciones del sistema de monitoreo:

#### 1. Eventos:

- Autorización para consultar.
- Finalizar consulta.
- Autorización para capturar.
- Finalizar captura.

#### 2. Condiciones:

- Personas consultando.
- Personas capturando.
- Personas inactivas.

Ahora analicemos la interacción entre los módulos del sistema. Como podemos observar, las condiciones que acabamos de señalar son los posibles estados del sistema y podemos considerarlos como los módulos del mismo. El proceso de consulta consiste en que la consulta se autorice, posteriormente se realice y finalmente ésta se concluya. El proceso de captura es análogo al de consulta. Ambos procesos tienen en común un estado: cuando no se consulta ni se captura, es decir, el estado de inactividad.

La representación gráfica de este sistema de monitoreo es la de la Figura 3.1, donde  $t_1$  representa la autorización para consultar;  $t_3$ , el fin de la consulta;  $t_2$ , la autorización para capturar; y  $t_4$ , el fin de la captura. Los lugares  $p_1$ ,  $p_2$  y  $p_3$  representan los estados del sistema: personas realizando consultas, inactivas y capturando, respectivamente. Los elementos en cada lugar indican la *cantidad* de personas. La mutua exclusión entre el capturista y los demás está señalada en los arcos que van de  $p_2$  a  $t_2$  y de  $t_4$  a  $p_2$  con el número 3. En esta red de Petri, el marcado inicial indica que el personal se encuentra en estado de inactividad.

Por último, cabe señalar que en esta red los eventos no ocurren en paralelo ya que no existe la posibilidad de disparar dos transiciones al mismo tiempo dadas las características del sistema. Si se deseara que esta red contara con transiciones en paralelo, bastaría con hacer que el peso de todos los arcos de la red fuera igual a 1, ya que de esta forma, y considerando que el estado inicial del sistema fuera el de inactividad, las transiciones  $t_1$  y  $t_2$  estarían activadas y, de desearse así, podrían dispararse simultáneamente. Sin embargo, es importante señalar que aun haciendo este pequeño cambio, la red no sería concurrente, pues para ello se requiere la *independencia* de las transiciones y en este caso las transiciones  $t_1$  y  $t_2$  comparten el *lugar de entrada*  $p_2$  y, por lo tanto, no son independientes.

*3.2 Ejemplo.* La descripción formal de la red de Petri de la Figura 3.1 es la siguiente:  $PN = \{P, T, I^-, I^+\}$  donde

- $P = \{p_1, p_2, p_3\}$ .
- $T = \{t_1, t_2, t_3, t_4\}$ .
- $I^-(p_1, t_3) = 1, I^-(p_2, t_1) = 1, I^-(p_2, t_2) = 3, I^-(p_3, t_4) = 1$ . Todos los demás valores de la función  $I^-$  son cero.
- $I^+(p_1, t_1) = 1, I^+(p_2, t_3) = 1, I^+(p_3, t_2) = 1, I^+(p_2, t_4) = 3$ . Todos los demás valores de la función  $I^+$  son cero.

*3.3 Ejemplo.* El marcado inicial de la red de Petri de la Figura 3.1 es

$$M_0(p_1) = 0$$

$$M_0(p_2) = 3$$

$$M_0(p_3) = 0$$

La entrada de elementos ya sea a lugares o bien a transiciones así como la salida de los mismos, se define formalmente como sigue:

**3.4 Definición.** Sea  $PN = (P, T, I^-, I^+)$  una red de Petri.

- Los lugares de entrada para la transición  $t$ :  $\bullet t := \{p \in P \mid I^-(p, t) > 0\}$ .
- Los lugares de salida para la transición  $t$ :  $t\bullet := \{p \in P \mid I^+(p, t) > 0\}$ .
- Las transiciones de entrada para el lugar  $p$ :  $\bullet p := \{t \in T \mid I^+(p, t) > 0\}$ .
- Las transiciones de salida para el lugar  $p$ :  $p\bullet := \{t \in T \mid I^-(p, t) > 0\}$ .

La extensión usual de conjuntos  $X \subseteq P \cup T$  se define como  $\bullet X = \bigcup_{x \in X} \bullet x$ ,  $X\bullet = \bigcup_{x \in X} x\bullet$ .

Esto quiere decir que si un arco va del lugar  $p$  a la transición  $t$ , decimos que  $p$  es un *lugar de entrada* para la transición  $t$ . Un *lugar de salida* se define de manera análoga.

Los conjuntos  $\bullet p$ ,  $\bullet t$  y  $p\bullet$ ,  $t\bullet$  son también conocidos, respectivamente, como preconjuntos y postconjuntos ya sea del lugar  $p$  o de la transición  $t$ , según sea el caso.

**3.4 Ejemplo.** En la red de Petri de la Figura 3.1:

- Las entradas de las transiciones son  $\bullet t_1 = \{p_2\}$ ,  $\bullet t_2 = \{p_2\}$ , es decir,  $\bullet\{t_1, t_2\} = \{p_2\}$ ,  $\bullet t_3 = \{p_1\}$ ,  $\bullet t_4 = \{p_3\}$ ; y las salidas,  $t_1\bullet = \{p_1\}$ ,  $t_2\bullet = \{p_3\}$ ,  $t_3\bullet = \{p_2\}$ ,  $t_4\bullet = \{p_2\}$ , o sea  $\{t_3, t_4\}\bullet = \{p_2\}$ .
- Las entradas de los lugares son  $\bullet p_1 = \{t_1\}$ ,  $\bullet p_2 = \{t_3, t_4\}$ ,  $\bullet p_3 = \{t_2\}$ ; y las salidas,  $p_1\bullet = \{t_3\}$ ,  $p_2\bullet = \{t_1, t_2\}$ ,  $p_3\bullet = \{t_4\}$ .

**3.5 Definición.** Sea  $PN = (P, T, I^-, I^+)$ .

1. A  $F \subseteq (P \times T) \cup (T \times P)$  dada por  $F := \{(x, y) \mid x, y \in P \cup T : x \in \bullet y\}$  se le llama *relación de flujo* de  $PN$ . Sea  $F^*$  la cerradura reflexiva y transitiva de  $F$ , i.e.,  $\forall x, y, z \in P \cup T$ :

- $(x, x) \in F^*$ .
- $(x, y) \in F \Rightarrow (x, y) \in F^*$ .
- $(x, y) \in F^*$  y  $(y, z) \in F^* \Rightarrow (x, z) \in F^*$ .

2.  $PN$  está débilmente conectada si y sólo si  $\forall x, y \in P \cup T : xF^*y$  o  $yF^*x$ .

3.  $PN$  está fuertemente conectada si y sólo si  $\forall x, y \in P \cup T : xF^*y$  y  $yF^*x$ .

Esto es:

1. Una red de Petri está débilmente conectada si desde cualquier lugar existe un camino (secuencia de transiciones) que conduzca a los demás lugares de la red, pero que no haga posible el regreso al lugar de origen.
2. Una red de Petri está fuertemente conectada si desde cualquier lugar existe un camino (secuencia de transiciones) que conduzca a *todos* los demás lugares de la red y que a su vez permita el regreso al lugar de origen.

El concepto de red de Petri fuertemente conectada es muy similar al de cadena irreducible estudiado previamente en el Capítulo 2, página 43.

### 3.3. Comportamiento Dinámico

El análisis del comportamiento de las redes de Petri permite calcular nuevos marcados a partir del marcado original y una serie de descargas y describe los cambios que sufren los estados del sistema dinámico de eventos discretos modelado a través de las redes de Petri.

El comportamiento dinámico está determinado por las activaciones y descargas de las transiciones como se explica a continuación.

**3.6 Definición.** Sea  $PN = (P, T, I^-, I^+, M_0)$  una red de Petri.

1. El *marcado* de la red de Petri es una función  $M : P \mapsto \mathbb{N}$ , donde  $M(p)$  denota el número de elementos en  $p$ . Esto es, el marcado de la red de Petri asigna a cada lugar un entero no negativo. Si el marcado asigna el entero no negativo  $n$  al lugar  $p$ , decimos que existen  $n$  elementos en  $p$ .
2. Se dice que el conjunto  $\tilde{P} \subseteq P$  es *marcado* en  $M$ , si y sólo si  $\exists p \in \tilde{P} : M(p) > 0$ ; de lo contrario, se dice que  $\tilde{P}$  es *no marcado* o *vacío* en  $M$ .
3. La transición  $t \in T$  está *activada* en  $M$ , si y sólo si  $M(p) \geq I^-(p, t)$ ,  $\forall p \in P$ , y se denota  $M[t >$ . Si cada lugar de entrada de una transición  $t$  tiene, al menos, el número de elementos señalados por la función de incidencia hacia atrás, decimos que  $t$  está activada.
4. La transición  $t \in T$ , activada en el marcado  $M$ , se *descarga* conduciendo a un nuevo marcado  $M'$  donde

$$M'(p) = M(p) - I^-(p, t) + I^+(p, t), \quad \forall p \in P,$$

lo cual se denota como  $M[t > M'$ . Decimos que  $M'$  es directamente alcanzable desde  $M$  y lo escribimos  $M \rightarrow M'$ . Sea  $\rightarrow^*$  la cerradura reflexiva y transitiva de  $\rightarrow$ . El marcado  $M'$  es alcanzable desde  $M$ , si y sólo si  $M \rightarrow^* M'$ .

5. La *secuencia de descargas* (secuencia de las ocurrencias) de una red de Petri es una secuencia finita de transiciones  $\sigma = t_1, \dots, t_n, n \geq 0$  tal que los marcados  $M_1, \dots, M_n$  satisfacen  $M_i[t_i > M_{i+1}, \forall i = 1, \dots, n$ . La secuencia de descargas vacía es denotada por  $\varepsilon$  y cumple siempre que  $M[\varepsilon > M$ .

El marcado inicial es el punto de partida del comportamiento dinámico de las redes de Petri y por ello se escribe simplemente como  $M_0$ . Como las transiciones sólo se activan si sus lugares de entrada tienen al menos tantos elementos como los establecidos por la función de incidencia hacia atrás, serán de especial interés los subconjuntos de lugares marcados.

Al modelar una situación, la descarga de una transición simula la ocurrencia de ese evento. Por supuesto, un evento puede ocurrir sólo si se cumplen todas las condiciones para su ejecución; es decir, la transición sólo se puede descargar si está activada.

La expresión  $M'(p) = M(p) - I^-(p, t) + I^+(p, t)$  corresponde a la idea intuitiva del proceso de descarga. Si la transición  $t$  se descarga, se destruirán  $I^-(p, t)$  elementos del lugar  $p$  y se crearán  $I^+(p, t)$  elementos en  $p$ . Note que el lugar  $p$  es el mismo en ambas expresiones,  $I^-(p, t)$  y  $I^+(p, t)$ .

En muchos casos alguna de las dos funciones de incidencia es igual a 0, pero existen algunos casos en los que una transición puede destruir y crear elementos en el mismo lugar a lo que suele llamársele *bucle*.

### 3.5 Ejemplo.

1. Ahora analicemos cuáles transiciones de la Figura 3.1 están activadas. En el caso de la transición  $t_1$  podemos observar que

$$\begin{aligned} M_0(p_1) &= I^-(p_1, t_1) \\ M_0(p_2) &> I^-(p_2, t_1) \\ M_0(p_3) &= I^-(p_3, t_1) \end{aligned}$$

La condición  $M(p) \geq I^-(p, t) \forall p \in P$  se cumple en el marcado  $M_0$  y por lo tanto  $M_0[t_1 >$ .

Haciendo este análisis para las demás transiciones, podemos comprobar que en el marcado  $M_0$  la transición  $t_2$  también está activada, mientras que las transiciones  $t_3$  y  $t_4$  no lo están.

2. Después de descargar la transición  $t_1$  se alcanza el marcado  $M_1$  desde el marcado inicial  $M_0$ , lo que se denota como

$$M_0[t_1 > M_1$$

Entonces

$$M_1 = M_0(p) - I^-(p, t_1) + I^+(p, t_1)$$

donde

$$\begin{aligned} M_1(p_1) &= M_0(p_1) + 1 = 1 \\ M_1(p_2) &= M_0(p_2) - 1 = 2 \\ M_1(p_3) &= M_0(p_3) = 0 \end{aligned}$$

3. Una posible secuencia de descargas para la Figura 3.1 es  $\sigma = t_1 t_1 t_3 t_3$ , la cual se ilustra en la Figura 3.2.

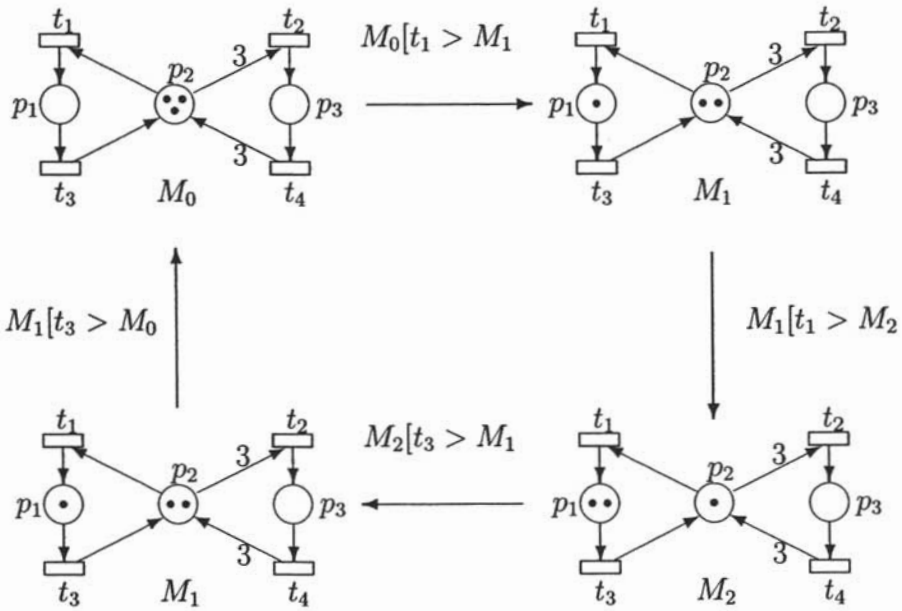
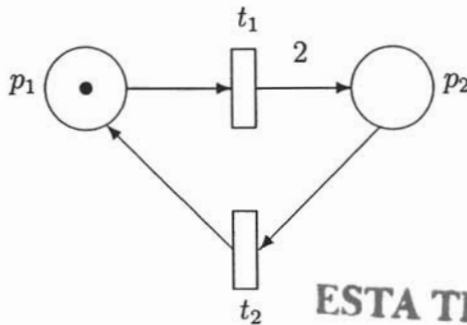


Figura 3.2: Secuencia de descargas  $\sigma = t_1t_1t_3t_3$



**ESTA TESIS NO SALE  
DE LA BIBLIOTECA**

Figura 3.3: Red de Petri no acotada



### 3.4. Propiedades

Una vez descrito el comportamiento dinámico de las redes de Petri nos ocuparemos de las propiedades de estas redes y en cómo verificarlas. Evidentemente, sólo estamos interesados en aquellas propiedades concernientes a los marcados alcanzables.

La importancia de dichas propiedades se desprende de la necesidad de demostrar la consistencia de las redes de Petri en el marcado inicial. En otras palabras, generalmente se requiere que los sistemas implementables modelados a través de las redes de Petri verifiquen dichas propiedades.

Para garantizar que no habrá un desbordamiento de elementos en los lugares, es necesario que exista una cota. Un marcado  $M$  para una red de Petri está *acotado* si existe algún entero positivo  $k$  con la propiedad de que, en cualquier secuencia de descarga, ningún lugar recibe más de  $k$  elementos. Si un marcado  $M$  está acotado y en cualquier secuencia de descargas ningún lugar recibe más de un elemento, decimos que  $M$  es un marcado *seguro*.

Otra propiedad es la *supervivencia* concerniente a las descargas de las transiciones. Un marcado  $M$  de una red de Petri está *vivo* si, partiendo de  $M$ , sin importar la serie de descargas realizadas, es posible descargar cualquier transición dada mediante alguna secuencia de descargas adicionales. Si un marcado  $M$  está vivo para una red de Petri  $PN$ , entonces, sin importar la serie de descargas de las transiciones,  $PN$  nunca se estancará. Como las transiciones modelan actividades, no resulta conveniente que el sistema llegue a un punto tal en el que no le sea posible activar nuevamente una transición. De ahí la importancia de esta propiedad.

Por último, es deseable que el conjunto de marcados alcanzables contenga exactamente un subconjunto de marcados fuertemente conectados. Esta noción es equivalente a la de conjunto irreducible. Sabemos gracias a la teoría markoviana estudiada en el capítulo anterior, que es necesario que el proceso objeto de estudio sea irreducible para que exista la distribución de estado estable. De manera similar cuando existe algún marcado  $M$  alcanzable desde todos los demás marcados, es decir, un conjunto alcanzable finito existe la distribución de estado estable.

Estas propiedades se resumen en la siguiente definición.

**3.7 Definición.** Sea  $PN = (P, T, I^-, I^+, M_0)$  una red de Petri.

1. El *conjunto alcanzable* de  $PN$  se denota  $R(PN) := \{M | M_0 \rightarrow^* M\}$ . Si  $PN$  denota una red de Petri no marcada o si queremos considerar a una parte del conjunto alcanzable, el conjunto de marcados alcanzables para un marcado  $\tilde{M}$  dado se denotará  $R(PN, \tilde{M}) := \{M | \tilde{M} \rightarrow^* M\}$ . Para una red de Petri marcada  $R(PN) = R(PN, M_0)$ .
2.  $PN$  es una red de Petri *acotada*, si y sólo si  $\forall p \in P : \exists k \in \mathbb{N} : \forall M \in R(PN) : M(p) \leq k$ .  $PN$  es *segura*, si y sólo si  $\forall p \in P : \forall M \in R(PN) : M(p) \leq 1$ .
3. La transición  $t \in T$  está *viva*, si y sólo si  $\forall M \in R(PN) : \exists M' \in R(PN) : M \rightarrow^* M'$  y  $M'[t >$ .  $PN$  está *viva*, si y sólo si todas sus transiciones están vivas, i.e.,  $\forall t \in T, M \in R(PN) : \exists M' \in R(PN) : M \rightarrow^* M'$  y  $M'[t >$ .
4. El marcado  $M \in R(PN)$  es un estado *hogar*, si y sólo si  $\forall M' \in R(PN) : M' \rightarrow^* M$ .

La seguridad es una propiedad de especial interés si los lugares de una red de Petri representan condiciones de tipo booleano, es decir, de tipo lógico: falso o verdadero. En tal caso la presencia o ausencia de un elemento en el lugar expresa la idea de que las condiciones se cumplen o no. Si, para algún marcado  $M$ , un lugar está vacío, se dice que la proposición asociada es falsa en el estado denotado por  $M$ . Si el lugar contiene un elemento, la proposición asociada se toma por cierta en el estado  $M$ .

Cuando modelamos condiciones booleanas estamos únicamente interesados en redes seguras, puesto que dos o más elementos en un lugar no tendrían sentido. Las redes *Evento-Condición* fueron de las primeras redes de Petri creadas y fueron especialmente diseñadas para este propósito. Dichas redes no permiten pesos de arcos múltiples y las transiciones se activan sólo cuando sus lugares de salida están vacíos, evitando así marcados con más de un elemento en un lugar. En este tipo de redes, al caso en el que los lugares de salida de una transición activada están marcados se le conoce como *contacto*.

Para simplificar un poco la notación ahora utilizaremos notación vectorial. Si el conjunto de lugares  $P$  está dado por  $\{p_1, \dots, p_n\}$ , la función  $M : P \rightarrow \mathbb{N}$  puede ser vista como un vector  $M := (M(p_1), \dots, M(p_n))^T$ .

**3.1 Teorema.** *Si  $PN$  está viva y acotada, entonces  $PN$  está fuertemente conectada.*

En otras palabras, una condición necesaria, pero no suficiente, para que una red de Petri esté viva y acotada es que esté fuertemente conectada. Como estamos interesados en redes de Petri que satisfagan estas condiciones, consideraremos únicamente redes de Petri fuertemente conectadas.

## 3.5. Análisis de las redes de Petri

Al modelar sistemas por medio de las redes de Petri frecuentemente se desea verificar las propiedades que las redes satisfacen. Existen dos métodos para este fin: el análisis del conjunto alcanzable y el análisis invariante. A continuación explicaremos brevemente en qué consisten.

### 3.5.1. Análisis del conjunto alcanzable

El *conjunto alcanzable* de una red de Petri se representa usualmente como un *árbol*, donde los nodos del árbol son los marcados de la red de Petri. Dos nodos  $M$  y  $M'$  se conectan con un arco dirigido si y sólo si  $M[t > M'$  para alguna  $t \in T$ . Este arco se etiqueta con  $t \in T$ .

El *árbol del conjunto alcanzable* se genera comenzando con el marcado inicial de la red de Petri y añadiendo los marcados directamente alcanzables a los cuales se les llama *hojas*. Luego seguimos con esos nuevos marcados y determinamos sus marcados directamente alcanzables. Esos marcados se convierten en nuevas hojas de la parte previamente generada del árbol del conjunto alcanzable, etc. Si alcanzamos un marcado previamente explorado ya no continuamos construyendo el árbol partiendo de ese nodo.

El árbol del conjunto alcanzable puede transformarse directamente en un grafo removiendo nodos múltiples y conectándolos apropiadamente. A este grafo se le conoce como *grafo del conjunto alcanzable*.

**3.8 Definición.** El *grafo del conjunto alcanzable* de la red de Petri cuyo conjunto alcanzable es  $R(PN, M_0)$  es el grafo en el cual:

- Los nodos son todos los elementos de  $R(PN, M_0)$ .
- Hay un arco orientado del marcado  $M$  al marcado  $M'$ , si y sólo si  $\exists t \in T : M[t > M'$ .

*3.6 Ejemplo.* Las Figuras 3.4 y 3.5 muestran el árbol del conjunto alcanzable y el grafo del conjunto alcanzable, respectivamente, correspondientes a la red de Petri de la Figura 3.1.

*3.7 Ejemplo.* A través del grafo del conjunto alcanzable (Figura 3.5) podemos apreciar todos los marcados alcanzables para cada uno de los lugares de la red de Petri de la Figura 3.1. Es claro que, considerando el marcado inicial, esta red está acotada y su cota es igual a 3. Asimismo puede verificarse que la red está viva, pues después de una cierta secuencia de descargas siempre es posible activar cada una de las transiciones que la conforman, lo cual quiere decir que ninguna de ellas queda permanentemente inhabilitada. Como la red está viva y es acotada, por el teorema 3.1, la red está fuertemente conectada.

Desafortunadamente, este método para crear el árbol del conjunto alcanzable falla en el caso de redes no acotadas. Sin embargo, existe un método alternativo para estos casos, el cual se estudia a detalle en el Capítulo 5 del libro de Bause & Kritzinger 1995 [3].

### 3.5.2. Análisis invariante

Un método alternativo para verificar las propiedades que las redes de Petri satisfacen es el análisis *p-invariante*, también conocido como el análisis de los lugares invariantes.

Hasta el momento hemos explicado cómo verificar que la sincronización del sistema modelado por la red de Petri de la Figura 3.1 sea consistente. Pero

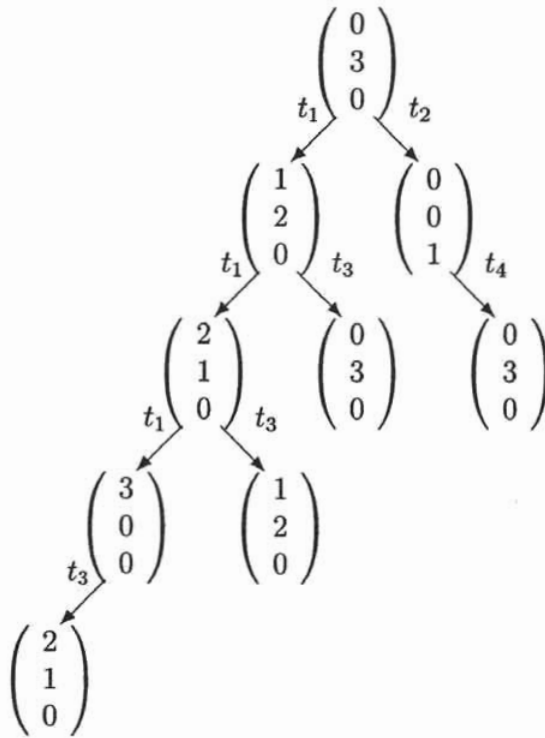


Figura 3.4: Árbol del conjunto alcanzable

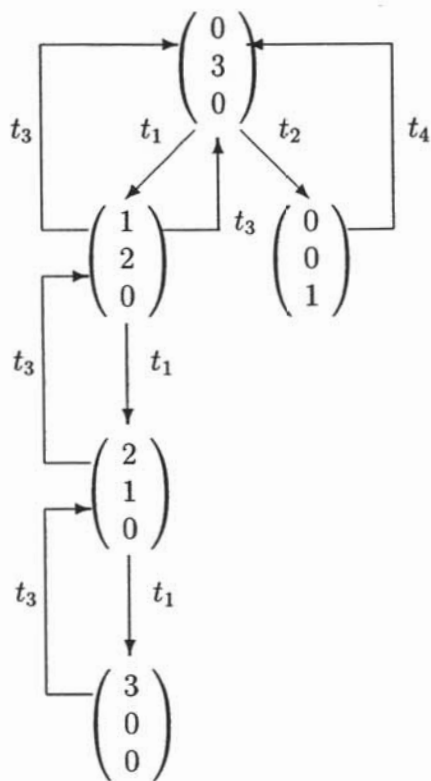


Figura 3.5: Grafo del conjunto alcanzable

no nos hemos ocupado de que los mecanismos de mutua exclusión funcionen adecuadamente. Comúnmente, para ello también se calculan los p-invariantes.

Como hemos señalado con anterioridad, un marcado puede ser representado con notación de funciones o con notación vectorial. Análogamente, las funciones de incidencia  $I^-$  e  $I^+$  pueden ser escritas como matrices. Dichas matrices ( $C^-$  y  $C^+$ ) reciben el nombre de *matrices de incidencia hacia atrás y hacia adelante*, respectivamente, y se definen como sigue.

**3.9 Definición.**  $PN = (P, T, I^-, I^+, M_0)$ .

La *matriz de incidencia hacia atrás*  $C^- = (c_{ij}^-) \in \mathbb{N}^{n \times m}$  se define como

$$c_{ij}^- := I^-(p_i, t_j), \forall p_i \in P, t_j \in T,$$

La *matriz de incidencia hacia adelante*  $C^+ = (c_{ij}^+) \in \mathbb{N}^{n \times m}$  se define como

$$c_{ij}^+ := I^+(p_i, t_j), \forall p_i \in P, t_j \in T,$$

y la *matriz de incidencia* para  $PN$  se define como  $C := C^+ - C^-$ .

Como podemos observar,  $C$  es una matriz de  $n$  renglones (lugares) y  $m$  columnas (transiciones) cuyos elementos pertenecen a  $\mathbb{N}$ .

Es importante señalar que la matriz de incidencia no es suficiente para determinar la estructura de una red de Petri, ya que para ello es necesario conocer además las matrices de incidencia hacia atrás y hacia adelante.

Las activaciones y descargas de las transiciones pueden expresarse entonces en términos de las matrices de incidencia. La transición  $t_i \in T$  está activada en el marcado  $M$ , si y sólo si  $M \geq C^- e_i$ , donde  $e_i$  es el  $i$ -ésimo vector unitario  $(0, \dots, \underbrace{1}_i, 0, \dots, 0)^T$ .

Si una transición activada  $t_i \in T$  se descarga en el marcado  $M$  dando como resultado el marcado  $M'$ , entonces escribimos

$$M' = M + C e_i \tag{3.1}$$

Esta ecuación recibe el nombre de *ecuación fundamental* de la red de Petri y establece que el proceso de descarga puede ser descrito sumando el vector

$Ce_i$  al marcado dado  $M$ , conduciendo al nuevo marcado  $M'$ .  $Ce_i$  es la  $i$ -ésima columna de la matriz de incidencia  $C$  y especifica la repercusión de  $t_i \in T$  en sus lugares de entrada y salida. Dado que los marcados pueden calcularse por medio de la adición de vectores, las redes de Petri pueden ser vistas como sistemas de sumas vectoriales con la restricción de que sólo transiciones activadas pueden descargarse.

Considere  $\sigma = t_{k_1} \dots t_{k_j}, j \in \mathbb{N}$  una secuencia de descargas con  $M_0[t_{k_1} > \dots M_{j-1}[t_{k_j} > M_j$ . El marcado  $M_j$  puede calcularse como sigue. Para cada marcado alcanzable la siguiente ecuación se cumple:

$$M_i = M_{i-1} + Ce_{k_i}, \quad i = 1, \dots, j$$

Sustituyendo los valores de  $i$  se tiene que

$$\begin{aligned} M_1 &= M_0 + Ce_{k_1} \\ M_2 &= M_1 + Ce_{k_2} = M_0 + C[e_{k_1} + e_{k_2}] \\ M_3 &= M_2 + Ce_{k_3} = M_0 + C[e_{k_1} + e_{k_2} + e_{k_3}] \\ &\vdots \end{aligned}$$

Lo que que conduce a la expresión

$$M_j = M_0 + C \sum_{i=1}^j e_{k_i}$$

Al vector  $f := \sum_{i=1}^j e_{k_i}$  se le conoce como *vector de descargas* o *vector característico de la secuencia de descargas*, y es de dimensión  $|T|$ . El componente  $f_i$  del vector  $f$  se refiere al número de ocurrencias de la transición  $t_i$  en  $\sigma$ .

**3.8 Ejemplo.** La Figura 3.1 tiene las siguientes matrices de incidencia:

$$C^+ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad C^- = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$C = \begin{matrix} & t_1 & t_2 & t_3 & t_4 \\ \begin{pmatrix} 1 & 0 & -1 & 0 \\ -1 & -3 & 1 & 3 \\ 0 & 1 & 0 & -1 \end{pmatrix} & p_1 \\ & & & & p_2 \\ & & & & p_3 \end{matrix}$$



1. En el marcado inicial,  $t_1$  está activada y el marcado resultante después de descargar  $t_1$  es

$$M' = M_0 + C \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$$

2. Supongamos la secuencia de descargas  $\sigma = t_1 t_1 t_3 t_3$ . Entonces, el vector de descargas es  $f = (2, 0, 2, 0)^T$  y con esta secuencia de descargas se alcanza el marcado

$$M'' = M_0 + C \begin{pmatrix} 2 \\ 0 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 0 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix}$$

lo cual puede apreciarse fácilmente en la Figura 3.2.

Debido a que un marcado alcanzable implica la existencia de una secuencia de descargas adecuada que conduce a dicho marcado, el siguiente teorema obviamente se cumple.

### 3.2 Teorema.

$$\forall M \in R(PN, M_0) : \exists f \in \mathbb{N}^m : M = M_0 + Cf \quad (3.2)$$

Sea  $v$  la función  $P \rightarrow \mathbb{Z}$ , es decir, la función que asocia un número entero a cada lugar  $p$  de la red de Petri.

**3.3 Teorema.** *Sea  $v$  la solución de la ecuación  $v^T C = 0$ , entonces la ecuación p-invariante se cumple:*

$$v^T M = v^T M_0, \quad \forall M \in R(PN, M_0) \quad (3.3)$$

donde  $v^T$  es el vector transpuesto de  $v$  y  $C$  es la matriz de incidencia de la red de Petri.

*Demostración.* Multiplicando  $v^T$  por ambos términos de la ecuación fundamental (3.1) se tiene que

$$v^T M = v^T M_0 + v^T C f$$

donde  $M$  es alcanzable mediante una serie de descargas  $\sigma$ . Si  $v^T C = 0$ , entonces

$$v^T M = v^T M_0$$

□

El teorema anterior es sumamente útil si se desea verificar las propiedades de todos los marcados alcanzables de una red de Petri, tal como se ilustra en el siguiente ejemplo.

*3.9 Ejemplo.* Consideremos nuevamente la red de Petri de la Figura 3.1. Como ya hemos visto, la matriz de incidencia de esta red es

$$C = \begin{array}{cccc} & t_1 & t_2 & t_3 & t_4 \\ \begin{pmatrix} 1 & 0 & -1 & 0 \\ -1 & -3 & 1 & 3 \\ 0 & 1 & 0 & -1 \end{pmatrix} & p_1 & p_2 & p_3 \end{array}$$

El vector

$$v = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix}$$

es una solución positiva de  $v^T C = 0$ . Como

$$v^T M = (1, 1, 3) \begin{pmatrix} M(p_1) \\ M(p_2) \\ M(p_3) \end{pmatrix} = M(p_1) + M(p_2) + 3M(p_3)$$

y

$$v^T M_0 = (1, 1, 3) \begin{pmatrix} 0 \\ 3 \\ 0 \end{pmatrix} = 3$$

De la ecuación (3.3) se obtiene que la ecuación p-invariante para esta red es

$$M(p_1) + M(p_2) + 3M(p_3) = 3 \tag{3.4}$$

El número de elementos en un lugar no puede ser negativo, por lo que de la ecuación ( 3.4) se derivan las siguientes restricciones

$$\forall M \in R(PN, M_0); M(p_1) \leq 3 \quad (3.5)$$

$$\forall M \in R(PN, M_0); M(p_3) \leq 1 \quad (3.6)$$

$$\forall M \in R(PN, M_0); M(p_1) + 3M(p_3) \leq 3 \quad (3.7)$$

De la ecuación ( 3.5) podemos concluir que  $M(p_1)$  varía de 0 a 3. Esto significa que es posible que hayan a lo más tres personas consultando la base de datos simultáneamente. De la ecuación ( 3.6) podemos concluir que en el lugar  $p_3$  el número de elementos puede ser igual a 0 ó 1. Esto indica la mutua exclusión entre capturistas ya que sólo puede haber uno en operación. La ecuación ( 3.7) prueba la mutua exclusión entre las personas que consultan y las que capturan porque si el número de personas que consultan no es cero, entonces necesariamente el número de capturistas es cero y viceversa. Así que el hecho de que la red de Petri de la Figura 3.1 representa las restricciones del sistema de monitoreo de Homobono ha sido probado sin necesidad de enumerar todos los marcados alcanzables a través del grafo del conjunto alcanzable.

Los invariantes  $v = (1, 0, 0)^T$  y  $v = (0, 0, 1)^T$  son otras dos posibles soluciones del sistema de ecuaciones lineales  $v^T C = 0$ . Sin embargo, no resultan relevantes al sustituirlos en la ecuación  $v^T M = v^T M_0$ , ya que simplemente expresan que el número de personas que se encuentra ya sea realizando consultas o capturando es constante, y de dichas ecuaciones no puede derivarse ninguna restricción.

El análisis de los *p-invariantes* frecuentemente se utiliza para verificar las propiedades de las redes de Petri. Para las redes de Petri vivas el siguiente teorema se cumple.

**3.4 Teorema.** *Sea  $PN$  una red de Petri viva y  $v \in \mathbb{Z}^n$ ,  $v \neq 0$ . Si  $v^T M = v^T M_0, \forall M \in R(PN, M_0)$ , entonces  $v$  es un  $p$ -invariante.*

*Demostración.* Como  $PN$  está viva, tenemos que  $\forall t_i \in T : \exists M, M' \in R(PN, M_0) : M[t_i > M'$  y  $M' = M + Ce_i$ . Multiplicando ambos miembros de esta ecuación por  $v^T$  obtenemos  $v^T M' = v^T M + v^T Ce_i$ .

Dado que  $v^T M = v^T M_0$ , entonces  $v^T Ce_i = 0$ , lo cual se satisface para todas las transiciones, es decir,  $\forall t_i \in T : v^T Ce_i = 0$ , y así  $v^T = 0$ , por lo que  $v$  es un *p-invariante*.  $\square$

Generalmente, la información proporcionada por los p-invariantes positivos es muy útil, ya que estos establecen una cota para todos los lugares cuyo número de elementos es diferente de cero y además no es necesario enumerar todos los marcados alcanzables.

Los p-invariantes positivos expresan la noción de conservación de un número limitado de elementos en cada lugar en todos los marcados alcanzables, por lo que es común que se empleen para verificar la existencia de cotas para la red de Petri. La base de todos los p-invariantes es el conjunto de restricciones que generalmente caracteriza a los marcados alcanzables.

**3.10 Definición.** *PN* está cubierta por p-invariantes positivos, si y sólo si  $\forall p_i \in P : \exists$  p-invariante  $v \in \mathbb{Z}^n$  con  $v \geq 0$  y  $v_i > 0$ .

**3.5 Teorema.** *Si PN está cubierta por p-invariantes positivos, entonces PN está acotada.*

*Demostración.* *PN* está cubierta por p-invariantes positivos  $\Rightarrow \exists$  p-invariantes  $v \in \mathbb{N}^n : v_i > 0, \forall i \in \{1, \dots, n\}$ . Con  $v^T M = v^T M_0 = k$  donde  $k$  es una constante,  $\forall M \in R(PN, M_0)$ , tenemos que  $v_i M(p_i) \leq v^T M_0 \forall p_i \in P$ , ya que  $M(p) \geq 0, \forall M \in R(PN, M_0)$ .

Si  $v_i > 0$ , entonces  $M(p_i) \leq \frac{v^T M_0}{v_i}$ , probando así que la red está acotada.  $\square$

Si la red no está cubierta por p-invariantes positivos, pero hay algunos lugares con p-invariantes positivos, entonces al menos esos lugares cubiertos están acotados.

**3.1 Corolario.** *Si existe el p-invariante  $v \in \mathbb{Z}^n : v \geq 0, v_i > 0$ , entonces  $p_i$  está acotado, i.e.,  $\exists k \in \mathbb{N} : \forall M \in R(PN, M_0) : M(p_i) \leq k$ .*

*Demostración.* La prueba se sigue del teorema 3.5.  $\square$

Otro invariante útil en el análisis de las redes de Petri es el *t-invariante* (transición invariante). Supongamos que tenemos un marcado  $M$  y que después de descargar algunas transiciones deseamos alcanzar el marcado  $M$  nuevamente.

Sea  $f \in \mathbb{N}^m$  el vector de descargas en cuestión. Si comenzamos en  $M$ , y alcanzamos a  $M$  nuevamente, siguiendo la ecuación (3.2), obtenemos la ecuación

$$M = M + Cf$$

lo cual implica que  $Cf = 0$ . Si deseamos regresar a algún marcado de la red de Petri entonces es necesario que el vector de descargas satisfaga  $Cf = 0$ . Sea  $w$  la función  $T \rightarrow \mathbb{Z}$ , es decir, la función que asocia un número entero a cada transición  $t$  de la red de Petri.

**3.11 Definición.**  $w \in \mathbb{Z}^m, w \neq 0$ , es un  $t$ -invariante si y sólo si  $Cw = 0$ .

**3.12 Definición.**  $PN$  está cubierta por  $t$ -invariantes positivos si y sólo si  $\forall t_i \in T : \exists$  el  $t$ -invariante  $w \in \mathbb{Z}^m$  con  $w \geq 0$  y  $w_i > 0$ .

*3.10 Ejemplo.* En el caso de la red de Petri de la Figura 3.1, existen dos posibles soluciones  $t$ -invariantes para el sistema de ecuaciones lineales  $Cw = 0$ :

$$w = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{y} \quad w = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

Estos dos  $t$ -invariantes pueden ser descargados desde el marcado inicial teniendo de esta forma las dos series de descargas  $\sigma = t_1 t_3$  y  $\sigma = t_2 t_4$ . Esto significa que cualquier comportamiento repetitivo del sistema puede obtenerse intercalando dos secuencias básicas:

- Autorización para consultar y finalizar consulta.
- Autorización para capturar y finalizar captura.

Esto demuestra la consistencia del sistema en cuanto a su mecanismo de sincronización.

Una condición necesaria para que la red de Petri esté viva y acotada es que esté cubierta por  $t$ -invariantes positivos.

**3.6 Teorema.**  $PN$  está viva y acotada  $\Rightarrow PN$  está cubierta por  $t$ -invariantes positivos.

*Demostración.* Al ser  $PN$  acotada su conjunto alcanzable es finito. Sea  $\tilde{R} \subseteq R(PN, M_0)$  un subconjunto fuertemente conectado terminal de  $R(PN, M_0)$ . Si escogemos un marcado arbitrario  $M \in \tilde{R}$  siempre podremos alcanzar a  $M$  después de hacer una serie de transiciones puesto que  $\tilde{R}$  está fuertemente conectada.

Si  $PN$  está viva entonces  $\exists f \in \mathbb{N}^m : f_i > 0, \forall i \in \{1, \dots, m\} : M = M + Cf$ , lo cual implica que  $Cf = 0$ .

Así,  $f$  es un  $t$ -invariante positivo que cubre a todas las transiciones de la red de Petri.  $\square$

**3.2 Corolario.** *Sea  $PN$  acotada. Si  $t_i \in T$  está viva entonces  $\exists$  el  $t$ -invariante  $w \in \mathbb{Z}^m : w \geq 0, w_i > 0$ .*

*Demostración.* La prueba se sigue del teorema 3.6.  $\square$

La complejidad del análisis invariante de las redes de Petri depende únicamente del número de lugares y transiciones que la componen y no del tamaño del conjunto alcanzable.

Al imponer ciertas restricciones a las redes de Petri surgieron diferentes clases de ellas y cada vez son más los investigadores que se han dado a la tarea de estudiarlas. En los libros de Bause & Kritzinger 1995 [3] y Cardoso & Camargo 1999 [8] se habla de algunas de las principales clases de redes de Petri que frecuentemente se utilizan para modelar sistemas modernos.



# Capítulo 4

## Redes de Petri estocásticas

*“Las construcciones de los matemáticos, como las de los pintores o los poetas, deben ser bellas; las ideas, como los colores o las palabras, deben encajar con armonía. La belleza es el primer requisito: no hay un lugar permanente en el mundo para las matemáticas feas”.*

G. H. Hardy

Las redes de Petri no involucran noción del tiempo, ya que no establecen en qué momento se activarán las transiciones. Es por esta razón que analizar el funcionamiento de un sistema a través de las redes de Petri que hemos estudiado hasta el momento no es posible. Las redes de Petri pueden ser utilizadas para hacer un análisis cualitativo del sistema, es decir, para analizar sus propiedades y verificar que sus mecanismos de sincronización y concurrencia sean los adecuados; en pocas palabras, para verificar la consistencia del sistema. Para poder realizar el análisis del funcionamiento, es decir, el análisis cuantitativo del sistema es necesario que el comportamiento del sistema a lo largo del tiempo forme parte de la descripción de la red de Petri.

Considerar el tiempo en el modelo es sumamente importante ya que en los sistemas reales casi siempre es indispensable considerarlo en la sincronización de los procesos.

A inicios de los años setenta surgieron múltiples propuestas para incorporar el



Tiempo Discreto	Tiempo Continuo
Distribución del tiempo de descarga/transición: Geométrica	Exponencial
Descargas/transiciones: Probabilidades $\mathbf{P}$	Tasas $\mathbf{Q}$
Solución de marcado/estado estable: $\Pi = \Pi \mathbf{P}, \sum \pi_i = 1$	$\Pi \mathbf{Q} = 0, \sum \pi_i = 1$

Figura 4.1: Redes de Petri estocásticas y teoría markoviana

factor tiempo a las redes de Petri. En general, existen dos posibilidades para hacer esto:

1. Asignándole duración a los lugares (nodos). Una vez que los elementos han sido descargados a un lugar  $p$ , éstos no estarán disponibles para activar ninguna transición sino hasta después de un periodo de tiempo.
2. Asignándole duración a las transiciones. Una vez que una transición ha sido activada ésta se disparará después de un periodo de tiempo.

En ambos casos, las redes de Petri pueden clasificarse, de acuerdo con la naturaleza del tiempo, como deterministas o estocásticas. Si el tiempo es determinista se les llama *redes de Petri temporales*; y si es aleatorio, *redes de Petri estocásticas*. Los pioneros en el estudio de las redes de Petri estocásticas fueron S. Natkin y M. K. Molloy, cada uno por su cuenta, a inicios de los años ochenta.

Resultan de particular interés las redes de Petri estocásticas en donde el tiempo tiene una distribución geométrica o bien exponencial, ya que estos casos pueden ser analizados a través de técnicas markovianas. La Figura 4.1 muestra la relación entre las redes de Petri estocásticas y la teoría markoviana. Particularmente, nos enfocaremos a estudiar el caso de las redes de Petri estocásticas donde el tiempo de descarga sigue una distribución exponencial.

**4.1 Definición.** La red de Petri estocástica de tiempo continuo  $SPN(PN, \Lambda)$

es aquella en la que  $PN = (P, T, I^-, I^+, M_0)$  y el conjunto  $\Lambda = (\lambda_1, \dots, \lambda_m)$ , donde  $\lambda_i$  es la tasa de descarga de la transición  $t_i$ .

Esto significa que el tiempo de descarga se distribuye exponencialmente, es decir, que la distribución acumulativa de la variable aleatoria  $\chi_i$ , la cual denota el tiempo de descarga de la transición  $t_i$ , es

$$F_{\chi_i}(x) = 1 - e^{-\lambda_i x}$$

Para poder hacer la analogía entre una red de Petri estocástica y un proceso markoviano es preciso obtener primero el grafo del conjunto alcanzable de la red de Petri. El espacio de estados del proceso de Markov será el conjunto alcanzable  $R(PN)$  y la tasa de transición del estado  $M_i$  al estado  $M_j$  estará dada por  $q_{ij} = \lambda_k$ . Es importante señalar que la tasa de descarga de la transición  $t_k$  de  $M_i$  a  $M_j$ ,  $\lambda_k$  puede ser dependiente de un determinado marcado.

Si varias transiciones van de  $M_i$  a  $M_j$  entonces  $q_{ij}$  será la suma de las tasas de descarga de esas transiciones. Si no existe ninguna transición de  $M_i$  a  $M_j$ , entonces  $q_{ij} = 0$  y  $q_{ii}$  se determina de forma que la ecuación  $\sum_j q_{ij} = 0$  se satisfaga.

La matriz cuadrada  $\mathbf{Q} = [q_{ij}]$  de orden  $s = |R(PN)|$  es la matriz  $\mathbf{Q}$  de la sección 2.2. Como ya hemos visto, la distribución de estado estable  $\Pi$  de los procesos de Markov se obtiene resolviendo el sistema de ecuaciones lineales

$$\Pi \mathbf{Q} = 0; \quad \sum_{i=1}^s \pi_i = 1$$

A partir del vector  $\Pi = (\pi_1, \pi_2, \dots, \pi_s)$  podemos calcular las siguientes medidas referentes al funcionamiento de la red de Petri estocástica de tiempo continuo.

**4.2 Definición.** Sea  $B \subseteq R(PN)$  un subconjunto de los marcados alcanzables de una red de Petri estocástica. Entonces, la probabilidad de estar en un estado elemento de dicho subconjunto está dada por

$$P[B] = \sum_{M_i \in B} \pi_i$$

**4.3 Definición.** Sea  $B(p_i, n)$  el subconjunto de  $R(PN)$  en el cual el número de elementos en el lugar  $p_i$  es  $n$ , i.e.  $B(p_i, n) = \{M \in R(PN) | M(p_i) = n\}$ . Entonces, el *número esperado de elementos* en el lugar  $p_i$  es

$$\bar{m}_i = \sum_{n=1}^{\infty} (nP[B(p_i, n)])$$

**4.4 Definición.** Sea  $EN_j$  un subconjunto de  $R(PN)$  en donde la transición  $t_j$  está activada, i.e.  $EN_j = \{M \in R(PN) | M[t_j >]\}$ . La *probabilidad de descarga de la transición  $t_j$*  es

$$r_j = \sum_{M_i \in EN_j} \pi_i \left( \frac{\lambda_j}{(-q_{ii})} \right)$$

donde  $(-q_{ii})$  es la suma de las tasas de transición hacia afuera de  $M_i$ .

*4.1 Ejemplo.* Analicemos la red de Petri estocástica de la Figura 4.2 (Falko & Bause 2002 [3]).

1. La transición  $t_1$  está activada en el marcado inicial  $M_0 = (1, 0, 0, 0, 0)^T$ . El tiempo de descarga se distribuye exponencialmente con tasa  $\lambda_1$ , i.e. el tiempo esperado de descarga de  $t_1$  es  $\frac{1}{\lambda_1}$ . Una vez que  $t_1$  se descarga, se obtiene el marcado  $M_1 = (0, 1, 1, 0, 0)^T$ . En  $M_1$ ,  $t_2$  y  $t_3$  están activadas de forma *concurrente*. Esto es, una de esas dos transiciones se descargará pasado un cierto periodo de tiempo. Si la transición  $t_2$  se descarga primero, obtendremos el marcado  $M_2 = (0, 0, 1, 1, 0)^T$  y si  $t_3$  se descarga antes que  $t_2$ , obtendremos el marcado  $M_3 = (0, 1, 0, 0, 1)^T$ . El marcado alcanzable a partir de  $M_1$  depende entonces de qué transición se activa primero. La probabilidad de que  $t_2$  se descargue primero<sup>1</sup> está dada por:

$$\begin{aligned} P[t_2 \text{ se descargue primero en } M_1] &= P[\chi_2 < \chi_3] \\ &= \int_0^{\infty} \left( \int_0^x \lambda_2 e^{-\lambda_2 y} dy \right) \lambda_3 e^{-\lambda_3 x} dx \\ &= \int_0^{\infty} (1 - e^{-\lambda_2 x}) \lambda_3 e^{-\lambda_3 x} dx \\ &= \frac{\lambda_2}{\lambda_2 + \lambda_3} \end{aligned}$$

---

<sup>1</sup>No hay que perder de vista que  $\chi_i$  denota el tiempo de descarga de la transición  $t_i$ .

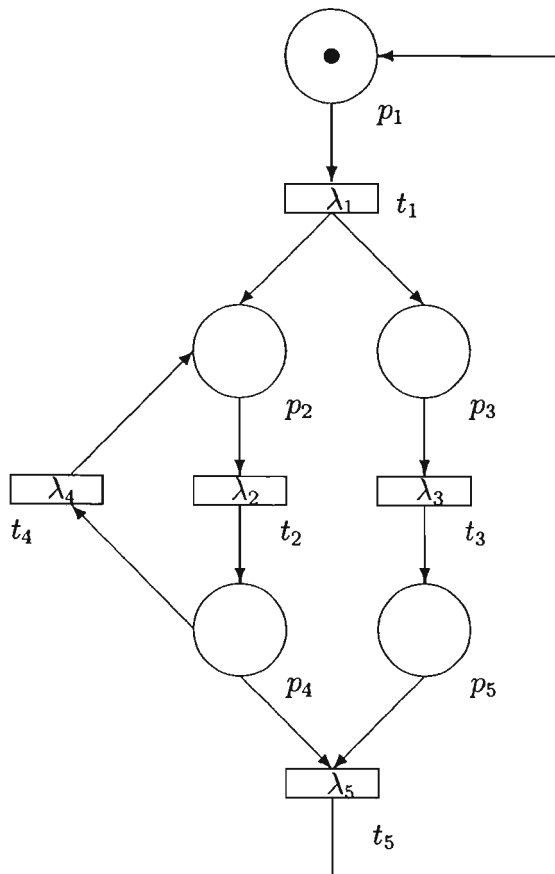


Figura 4.2: Red de Petri estocástica

y de manera similar

$$P[t_3 \text{ se descargue primero en } M_1] = \frac{\lambda_3}{\lambda_2 + \lambda_3}$$

Esto prueba que la probabilidad de alcanzar cualquier otro marcado a partir de  $M_1$  es independiente del tiempo que se permanezca en  $M_1$ .

2. El tiempo que se permanece en  $M_1$  está dado por el mínimo de los tiempos de descarga independientes, los cuales se distribuyen exponencialmente:

$$\begin{aligned} P[\min(\chi_2, \chi_3) \leq x] &= P[\chi_2 \leq x \text{ o } \chi_3 \leq x] \\ &= 1 - P[\chi_2 > x \text{ y } \chi_3 > x] \\ &= 1 - e^{-\lambda_2 x} e^{-\lambda_3 x} \\ &= 1 - e^{-(\lambda_2 + \lambda_3)x} \end{aligned}$$

Y de esta forma concluimos que el tiempo de permanencia en  $M_1$  se distribuye exponencialmente con parámetro  $(\lambda_2 + \lambda_3)$ .

*4.2 Ejemplo.* La Figura 4.4 muestra el proceso de Markov correspondiente a la red de Petri estocástica de la Figura 4.2. Asumamos las siguientes tasas de descarga:  $\lambda_1 = 2$ ,  $\lambda_2 = 1$ ,  $\lambda_3 = 1$ ,  $\lambda_4 = 3$ , y  $\lambda_5 = 2$ .

Comenzando con el marcado inicial en el que existe un elemento en el lugar  $p_1$  y ninguno en los demás lugares de la red, el conjunto alcanzable tiene cinco marcados o equivalentemente, el proceso de Markov cuenta con cinco estados.

La matriz de tasas de transición es

$$\mathbf{Q} = \begin{array}{c} M_0 \\ M_1 \\ M_2 \\ M_3 \\ M_4 \end{array} \begin{array}{ccccc} M_0 & M_1 & M_2 & M_3 & M_4 \\ \left( \begin{array}{ccccc} -2 & 2 & 0 & 0 & 0 \\ 0 & -2 & 1 & 1 & 0 \\ 0 & 3 & -4 & 0 & 1 \\ 0 & 0 & 0 & -1 & 1 \\ 2 & 0 & 0 & 3 & -5 \end{array} \right) \end{array}$$

Y resolviendo el proceso de Markov partiendo de las ecuaciones  $\Pi \mathbf{Q} = 0$  y

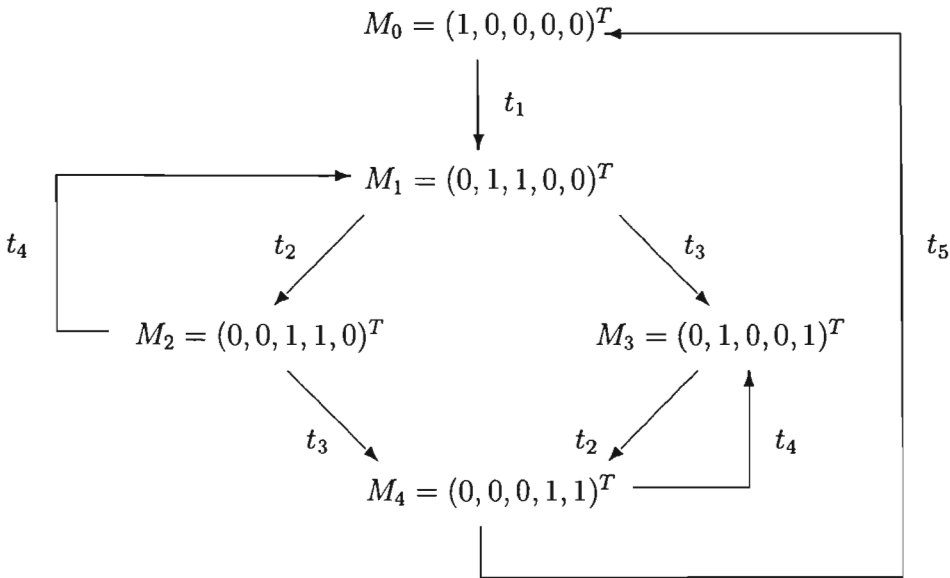


Figura 4.3: Grafo del conjunto alcanzable de la red de Petri estocástica

$\sum_{i=1}^s \pi_j = 1$ , obtenemos las siguientes probabilidades de marcados estables:

$$P[M_1] = P[(1, 0, 0, 0, 0)^T] = \frac{5}{43}$$

$$P[M_2] = P[(0, 1, 1, 0, 0)^T] = \frac{8}{43}$$

$$P[M_3] = P[(0, 0, 1, 1, 0)^T] = \frac{2}{43}$$

$$P[M_4] = P[(0, 1, 0, 0, 1)^T] = \frac{23}{43}$$

$$P[M_5] = P[(0, 0, 0, 1, 1)^T] = \frac{5}{43}$$

Apartir de las probabilidades de marcado estable y del número de elementos en cada lugar en un marcado particular podemos calcular las probabilidades

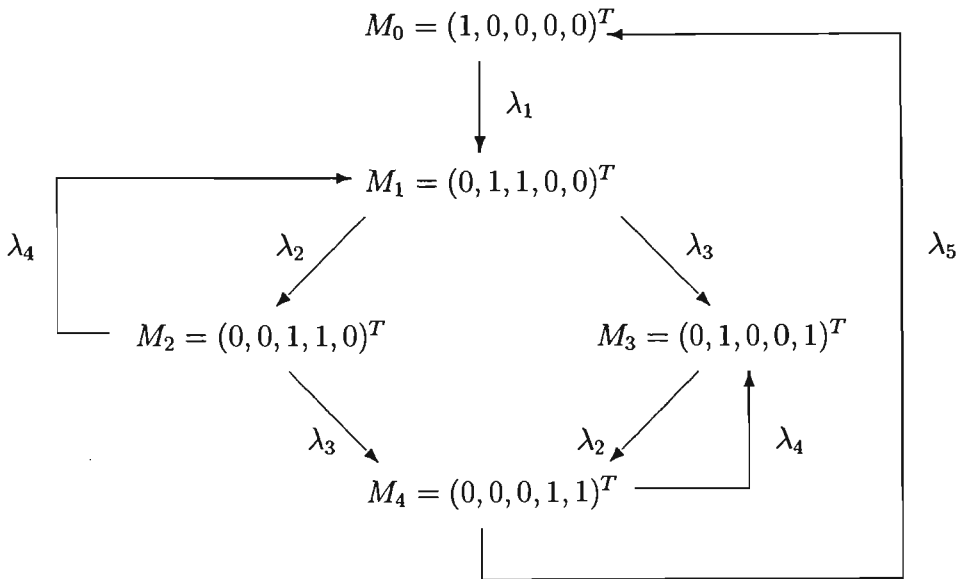


Figura 4.4: Proceso de Markov de la red de Petri estocástica

de que en el largo plazo existan  $\mu_i$  elementos en el lugar  $p_i$ .

$$P[\mu_1 = 0] = \frac{38}{43} \quad P[\mu_1 = 1] = \frac{5}{43}$$

$$P[\mu_2 = 0] = \frac{12}{43} \quad P[\mu_2 = 1] = \frac{31}{43}$$

$$P[\mu_3 = 0] = \frac{33}{43} \quad P[\mu_3 = 1] = \frac{10}{43}$$

$$P[\mu_4 = 0] = \frac{36}{43} \quad P[\mu_4 = 1] = \frac{7}{43}$$

$$P[\mu_5 = 0] = \frac{15}{43} \quad P[\mu_5 = 1] = \frac{28}{43}$$

Note que si asumimos un marcado inicial distinto obtendremos un proceso markoviano distinto y que mientras más elementos existan en el sistema el proceso markoviano será más extenso.

El aspecto más importante de las redes de Petri estocásticas es que el grafo del conjunto alcanzable de las redes de Petri y las cadenas, o procesos según sea

el caso, de Markov son isomórficos. En otras palabras, el número de marcados/estados y la estructura de los arcos que los conectan son los mismos tanto para las redes de Petri como para las cadenas de Markov. Las cadenas de Markov describen el árbol del conjunto alcanzable de la red de Petri estocástica. Para la demostración formal de la analogía entre la teoría markoviana y las redes de Petri se recomienda la publicación de Morales, Navarro & Pérez 1999 [24].

Todas las propiedades de las redes de Petri ordinarias son las mismas para las redes de Petri estocásticas. Por esta razón, el análisis cualitativo de las redes de Petri estocásticas puede realizarse con las mismas técnicas utilizadas para las redes de Petri ordinarias.

## 4.1. Posibles aplicaciones

En general, las redes de Petri estocásticas permiten no sólo evaluar la correcta estructura de un sistema sino además el correcto funcionamiento del mismo bajo un enfoque aleatorio.

En virtud de la analogía entre redes de Petri estocásticas y la teoría markoviana, prácticamente cualquier sistema que pueda ser modelado a través de la teoría markoviana puede serlo también a través de las redes de Petri estocásticas. La importancia de las redes de Petri estocásticas radica en este aspecto, pues éstas permiten construir modelos markovianos que además incorporan mecanismos de sincronización, concurrencia y/o mutua exclusión, por lo que pueden ser más cercanos al sistema en cuestión.

A continuación se presentan algunas posibles aplicaciones de las redes de Petri estocásticas.

### 4.1.1. Modelos de crecimiento poblacional

El tamaño y composición de una población, cualquiera que esta sea, está fluctuando constantemente y, generalmente, en forma aleatoria. Algunos fenóme-



nos biológicos que pueden ser modelados a través de redes de Petri estocásticas son:

- (i) Las consecuencias de mutaciones genéticas y de recombinaciones de genes en la teoría de la evolución.
- (ii) La distribución espacial de plantas y de comunidades animales.
- (iii) La lucha por la supervivencia entre dos poblaciones que interactúan o compiten.
- (iv) La propagación de epidemias.

#### **4.1.2. Comunicación y control**

Una amplia gama de problemas que involucran comunicación y/o control (como la recepción de señales de radio en presencia de disturbios naturales o artificiales, la reproducción de sonidos o imágenes, el diseño de sistemas de control para procesos industriales, las proyecciones, el análisis de fluctuaciones económicas y el análisis de registros que representan observaciones a lo largo del tiempo) son frecuentemente modelados a través de procesos estocásticos para evaluar los sistemas de comunicación y control en términos de su comportamiento promedio sobre un rango de circunstancias descritas probabilísticamente.

Para el diseño de sistemas de comunicación y de control óptimos el análisis de la interacción entre los módulos que los conforman es esencial y es por ello que las redes de Petri estocásticas podrían ser de gran utilidad en estas áreas.

#### **4.1.3. Administración e Investigación de Operaciones**

Las redes de Petri estocásticas proveen un método para el estudio cualitativo y cuantitativo de operaciones administrativas y, en consecuencia, son de vital importancia en la ciencia administrativa y la investigación de operaciones.

Dos áreas en las cuales los procesos estocásticos han tenido mayor aplicación son el control de inventarios y el análisis de líneas de espera, también conocido como teoría de colas.

**Control de Inventarios.** Para diversas organizaciones, como tiendas de mayorero o menudeo, almacenes, distribuidores de mercancías, manufactureras, entre otras, dos problemas de considerable importancia son:

- (i) Decidir cuando surtirse nuevamente de productos.
- (ii) Decidir la cantidad de productos con la que se surtirán.

El propósito del control de inventarios es minimizar el costo de mantenimiento del inventario y al mismo tiempo mantener un número de existencias suficiente para satisfacer la demanda aleatoria y para contar con una reserva en caso de que la entrega de productos por parte de algún proveedor se retrase.

Las fluctuaciones de inventarios pueden ser modeladas a través de procesos estocásticos eficientemente. Si incorporamos a nuestro modelo factores como la sincronización, la concurrencia o bien la mutua exclusión, podríamos hacerlo más apegado a la realidad. De ahí que las redes de Petri estocásticas sean herramientas potencialmente útiles en este campo.

**Teoría de colas.** Una cola (o línea de espera) se genera cuando clientes (o servidores) que llegan a algún lugar a recibir (o prestar) un servicio deben esperar para recibir (o prestar) dicho servicio. El grupo que espera para recibir (o prestar) un servicio, incluyendo a aquellos que están recibiendo (o prestando) el servicio, es llamado cola.

Existen muchos ejemplos de colas. Personas que esperan en las taquillas del cine, en las ventanillas del banco o en las cajas del supermercado forman una cola así como las personas que esperan en la estación de autobuses o en el aeropuerto ya sea para comprar boletos o para abordar. Aviones que aterrizan en un aeropuerto forman una cola. Los barcos que llegan a puertos a descargar y cargar nuevamente mercancías hacen una cola.

En la teoría de colas, las líneas de espera son clasificadas matemáticamente de acuerdo con cuatro aspectos:

- (i) La ley de probabilidad, o la distribución, que rige el tiempo entre llegadas sucesivas de consumidores o clientes.
- (ii) La ley de probabilidad, o la distribución, que rige el tiempo que toma atender a cada consumidor o cliente.
- (iii) El número de servidores.
- (iv) La manera en que los consumidores o clientes serán seleccionados para prestarles el servicio; posibles políticas de atención son: Primero en llegar, primero en salir (FIFO por sus siglas en inglés *First In, First Out*), selección de atención aleatoria, atención por orden de prioridad.

La teoría de colas estudia el efecto que tienen estos cuatro aspectos en varias cantidades de interés, tales como la longitud de la cola y el tiempo de espera de los consumidores para recibir el servicio.

Puesto que en las colas la sincronización y concurrencia de eventos juegan un papel importante, Bause & Kritzinger 1995 [3] ya han propuesto una extensión de las redes de Petri estocásticas aplicadas a la teoría de colas, a la que han llamado *Queuing Petri Nets*.

#### 4.1.4. Simulación

Las redes de Petri estocásticas son formalismos que brindan información del comportamiento dinámico del sistema modelado, esto es, además de modelar la estructura de un sistema permiten analizar su comportamiento, y llevar al modelo a condiciones límite, que en un sistema real son difíciles de lograr o sumamente costosas. Para llevar al modelo a dichas condiciones límite es conveniente hacer uso de la simulación.

En internet existen varios programas de redes de Petri estocásticas que permiten simular la ocurrencia de los eventos que conforman al sistema y visualizar la representación gráfica del mismo, así como el comportamiento de la red a lo largo del tiempo. Algunos sitios de interés son <http://www.informatik.uni-hamburg.de/TGI/PetriNets/tools/java/> en donde se ofrecen varios vínculos a programas que simulan el funcionamiento de las redes de Petri y la página

<http://pdv.cs.tu-berlin.de/~timenet>, en la cual se puede encontrar información sobre el paquete TimeNET (*Timed Net Evaluation Tool*) que modela sistemas a través de redes de Petri estocásticas. Véase también el artículo de Morales, Navarro & Pérez 1999 [24] titulado *LENA: Un Lenguaje para Especificar Redes de Petri Estocásticas*.

Podemos concluir señalando algunas de las principales ventajas de las redes de Petri estocásticas:

- Son herramientas formales que permiten modelar sistemas y ofrecen una representación gráfica de procesos.
- Son fáciles de entender.
- Son flexibles.
- Pueden ser analizadas (o simuladas) por computadora.
- Existe una importante analogía entre las redes de Petri estocásticas y la teoría markoviana.

Con esto daremos por terminado nuestro estudio sobre redes de Petri. En el libro de *Lecture Notes in Computer Science 2000* [20] se encuentra más información sobre teoría y aplicaciones de redes de Petri y para ahondar en el tema de redes de Petri estocásticas se sugiere el libro de Lindemann 1998 [21].



# Conclusiones y recomendaciones

A lo largo de la presente Tesis se han presentado, explicado y vinculado los conceptos fundamentales en materia de teoría markoviana y redes de Petri estocásticas como base de la teoría de la modelación matemática de sistemas empleando una notación accesible e incluyendo más de 20 ejemplos, de los cuales 18 son inéditos.

Este trabajo presenta a las redes de Petri no sólo como herramientas formales que permiten el análisis cualitativo y cuantitativo de sistemas, sino además como métodos alternativos para el estudio de sistemas que involucren aleatoriedad, logrando así el primer paso para iniciar una línea de investigación dirigida a la aplicación de redes de Petri en la ciencia actuarial.

Adicionalmente, la definición de probabilidad se ha construido partiendo de los conceptos de álgebra booleana y álgebra booleana de conjuntos y se ha explicado que el concepto de variable aleatoria se desprende también del álgebra booleana.

De este modo se han alcanzado los objetivos planteados al inicio de este trabajo.

Por otra parte, las hipótesis fueron sustentadas en ejemplos:

Se ha presentado a las redes de Petri estocásticas como producto de la conjunción de las teorías markoviana y de redes de Petri y, adicionalmente, se han propuesto algunas aplicaciones de las redes de Petri estocásticas en áreas en donde el actuario tiene oportunidades de desarrollo potenciales, por lo cual la conjunción de dichas teorías son benéficas para la carrera de Actuaría.

A través de la teoría markoviana puede llevarse a cabo el análisis cuantitativo de las redes de Petri al incorporar al modelo técnicas que permiten el estudio del comportamiento aleatorio del sistema, concretamente, cuando el tiempo de descarga de las transiciones que lo componen tiene una distribución ya sea geométrica o exponencial. Y las redes de Petri permiten el análisis cualitativo de las cadenas y procesos markovianos al incorporar al modelo las condiciones y eventos que conducen a los diferentes estados que los conforman. Por lo que podemos concluir que la teoría markoviana enriquece a la teoría de las redes de Petri, y viceversa.

En este proyecto se empleó como editor de textos el lenguaje  $\text{\LaTeX} 2_{\epsilon}$  debido a la gran cantidad de simbología matemática empleada, ya que a pesar de su complejidad es empleado por la calidad de su tipografía por las más importantes casas editoriales. Asimismo, en los principales congresos internacionales y revistas científicas arbitradas se exige la presentación de proyectos en  $\text{\LaTeX} 2_{\epsilon}$ , razón por la cual una de las conclusiones de este proyecto es el impulso a la utilización de este lenguaje por los estudiantes de las áreas de matemáticas para el desarrollo de sus trabajos de investigación, Tesis de Licenciatura y de Posgrado.

En el presente, el campo profesional del actuario ha evolucionado de forma notable, y se ha extendido a campos que inicialmente no eran considerados tradicionales dentro de la Actuaría. Entre las diversas áreas para ejercer dicha profesión se encuentran la industria, las ciencias médicas y biológicas, la planeación, las ciencias administrativas y económicas, entre muchas otras, además del campo clásico concerniente a los seguros, las finanzas y la estadística.

El actuario moderno recibe una fuerte formación matemática, que le permite manejar herramientas actuales para el buen ejercicio de su profesión. Una de estas herramientas, la cual ha sido poco difundida, la constituyen las redes de Petri estocásticas y su relación con los procesos markovianos. En este sentido la presente Tesis otorga elementos que acercan al actuario a estas herramientas de gran eficacia, pero que han sido poco difundidas y escasamente aplicadas. Este es un primer paso para el consecuente uso de dichas técnicas en la resolución de diversos problemas que enfrenta nuestra nación, lo cual constituye el reto para las presentes y futuras generaciones de actuarios.

# Apéndice





UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO  
FACULTAD DE ESTUDIOS SUPERIORES CUAUTITLAN

THE FIRST INTERNATIONAL CONFERENCE ON NONLINEAR, FRACTIONAL AND STOCHASTIC DYNAMICS AND THE FIRST INTERNATIONAL WORKSHOP ON COMPLEX SYSTEMS, APPLIED CATEGORY THEORY, FRACTALS, FRACTIONAL CALCULUS AND CHAOS

This is to certify that the following invited talk

*Stochastic Processes and Operads of Graphs*

by

*Mayra Lorena Diaz Sosa*

Was presented at the Facultad de Estudios Superiores Cuautitlan

January 24 to 28, 2005.

Profra. Soledad Carreto-G.  
Organizer of the International Conference

Dr. Juan A. Monjaraz-Crespo  
Director, FES-UNAM

Prof. Guillermo A. Vázquez-C.  
Organizer of the International Workshop



# Some Basic Notions of Probability and Markov Chains<sup>\*†</sup>

Mayra Lorena Díaz Sosa

Universidad Nacional Autónoma de México  
Facultad de Estudios Superiores Acatlán  
Jefatura del Programa de Actuaría.  
Av. Alcanfores y San Juan Totoltepec s/n, Col. Santa Cruz Acatlán,  
C.P. 53150, Naucalpan de Juárez, Estado de México.  
malodiso@exalumno.unam.mx

**Date:** February 28, 2005.

**Abstract.** This paper states some of the most important definitions about probability and stochastic theory, focusing particularly in the stochastic processes known as Markovian chains. Since there is a lack of a material that provides a logic order which allows us to understand the concept of probability, one of the achievements of this paper is stating a definition of probability through Boolean algebra and Boolean algebra of sets. Moreover, it establishes some basic definitions with a notation that makes possible an easy understanding of the meaning of the probabilistic and stochastic concepts, thus implying an additional achievement.

---

\*Lecture during “The First International Conference on Nonlinear, Fractional and Stochastic Dynamics”, January 24-28, 2005. Facultad de Estudios Superiores Cuautitlán.

†Fragment of the thesis “Stochastic Processes and Stochastic Petri Nets” (in process). Supported by UNAM, Dirección General de Asuntos del Personal Académico, Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), IN-105402 and by Consejo Nacional de Ciencia y Tecnología (CONACyT) de México, grant U 41214 F.

# 1 Introduction

The existence of randomness has been long discussed, and stating that the universe is random or deterministic is a very risky issue. Modeling a subsystem of the real world through a random or deterministic mathematical system is more useful and less ambitious. In this case we will only consider a subsystem of the universe as a deterministic one if, ignoring unexpected external effects, every time we turn it back to what seemingly is the same initial state, it produces the same effect during a nonnull period of time. We will consider it otherwise random if undistinguishable original states can immediately produce very different results. Once established the above-mentioned, we will be able to decide if a deterministic or a random model can be adopted in order to study the subsystem of the universe.

A mathematical model is a set of equations which purpose is to describe the behavior of a system. We can only assure whether a model is random or deterministic when comparing observations of the real world with a concrete model.

Stochastic theory is the mathematical theory supporting random models. This theory is increasingly stronger due to the fact that it allows building models close to reality. Random models however have not been exploded properly because of the calculation complexity they involve. In this article, we will refer to random models as stochastic processes.

Stochastic processes are very interesting for actuaries, the specialists in the evaluation, management and prevention of risks in insurance, finance, statistics, demography, and systems among others.

For example, market behavior, main economic indicators' behaviors, primes an assurance company has to collect from its clients and reserves it must have in order to meet its obligations, project evaluation, and resources or time optimization are events that involve risks the actuary, due to his scope, is usually interested in analyzing.

An actuary assumes the risk as a stochastic system, using therefore the stochastic processes as a mathematical tool for its study.

In order to understand the concept of stochastic process, we must first define the concept of probability. However, probability can not be understood without Boolean algebra. While reviewing literature about stochastic processes I realized that most of them do not even mention Boolean algebra. Some authors, like Durrett [1], Papoulis [3], Parzen [4] and Ross [5], make a brief review of probability as an introduction to stochastic processes, but do not

consider Boolean algebra. Many other authors, like Vélez [7], do not even make a review of probability. Due to the aforementioned and to the lack of material that provides a logic order which allows us to understand the concept of probability, one of the achievements of this paper is to stating a definition of probability under a Boolean algebra approach. Sometimes the notation used in several mathematical books is difficult to understand which might cause confusion. This article states some basic definitions of the stochastic and the probability theories with a notation that allows us to understand easily the meaning of the concepts, which implies another achievement.

## 2 Probability review

Understanding the stochastic theory requires the acknowledgement of some probability basic concepts. The following is a brief review.

**Definition 1.** A Set function  $f$  is defined by  $f : Q \rightarrow \mathbb{R}$ , where  $f$  is a set function if  $Q$  is a family of sets and  $\mathbb{R}$  is a subset of the real numbers.

**Definition 2.** A set function is of finite additivity if for each pair of elements  $A$  and  $B$ , of the family, disjoint and which union is element of  $Q$ , satisfies that  $f(A \cup B) = f(A) + f(B)$ . A set function with finite additivity, defined nonnegative is called measure with finite additivity or simply, measure.

**Definition 3.** Boolean algebra is the set  $\mathcal{B}$  with two binary operations  $\wedge$ ,  $\vee$ , the complement  $'$  and two constants,  $0$  and  $1$ , that satisfy the commutative, associative, idempotent, absorption, Morgan, identity and complement laws.

**Definition 4.** Boolean algebra of sets defines all the operations  $W$  of the universal set  $U$  according to the following rules:  $x \vee y = x \cup y$ ,  $x \wedge y = x \cap y$ ,  $x' = U \setminus x$ ,  $1 = U$ ,  $0 = \emptyset$ .

From these definitions, it is easy to prove that if  $f$  is a set function with finite additivity defined over a Boole algebra of sets  $Q$ , and if  $A$  and  $B$  are any element of  $Q$ , we obtain  $f(A \cup B) = f(A) + f(B) - f(A \cap B)$

**Definition 5.** Sample space associated to a random experiment, is the set of all possible results of such experiment. Each element of the sample space  $S$  is called elementary success, and the elements of a Boole algebra of sets of the sample space  $S$  are simply successes or events.

If two events  $A$  and  $B$  are such that  $A \cap B = \emptyset$ , then both events are *mutually exclusive* or disjoint. With all the elements previously mentioned we can now build in an axiomatic form the concept of probability.

**Definition 6.** Let  $S$  be a finite sample space and  $Q$  a Boole algebra whose elements are a subset of  $S$ . The function  $P$  defined in  $Q$  is called probability if it satisfies:

1.  $P(A) \geq 0$
2.  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3.  $P(S) = 1$

So the probability space is every tern  $(S, Q, P)$ .

The *exhaustive set of events* is an important concept required to understand the stochastic theory. The union of this set of events creates the sample space  $S$  of all possible outcomes of an experiment. *Mutually exclusive events* are the events  $\{A_1, A_2, \dots, A_n\}$  such that  $A_i A_j = A_i \cap A_j = \emptyset$  for all  $i \neq j$  and  $A_1 \cup A_2 \cup \dots \cup A_n = S$ .

The *conditional probability* of an event  $A$ , given that event  $B$  occurred is denoted by  $P[A|B]$  and defined as

$$P[A|B] := \frac{P[AB]}{P[B]}$$

whenever  $P[B] \neq 0$ . The *statistical independence* of events can be defined as follows. Two events  $A$  and  $B$  are said to be statistically independent if, and only if,  $P[AB] = P[A]P[B]$ . Moreover, for two independent events  $A$  and  $B$   $P[A|B] = P[A]$ , which means that the occurrence of an event  $B$  does not affect the occurrence probability of the independent event  $A$  in any way and vice versa.

**Theorem 1.** Theorem of total probability. Consider an event  $B$  and the set of mutually exclusive events  $\{A_1, A_2, \dots, A_n\}$ . If the event  $B$  is to occur, it must be in conjunction with exactly one of the mutually exclusive events  $A_i$ . That is  $P[B] = \sum_{i=1}^n P[A_i B]$

From the definition of conditional probability we can obtain a second form of the theorem of total probability:  $P[B] = \sum_{i=1}^n P[B|A_i]P[A_i]$ . With all these concepts in mind we can understand the *Bayes' theorem*.

**Theorem 2.** Bayes' theorem. *Let  $\{A_i\}$  be a set of mutually exclusive and exhaustive events. Then*

$$P[A_i|B] = \frac{P[B|A_i]P[A_i]}{\sum_{j=1}^n P[B|A_j]P[A_j]}$$

This theorem allows calculating the probability of a conditioned event on a second by calculating the probability of the second conditioned on the first one as well as on other terms.

Let us assume that we have an experiment  $F$  whose possible outcomes  $\zeta$  are various objects that match with certain event  $S$ . Now, according to some rule, lets assign to each  $\zeta$  a number  $\chi(\zeta)$ . The relationship between the elements  $\zeta$  of the set  $S$  and certain numbers is known as random variable and it is denoted by  $\chi$ .

In other words, the random variable  $\chi$  is the function whose domain is the set  $S$  of all possible outcomes of an experiment. The possible outcomes of an experiment can be objects identified during the performance of the experiment.  $\zeta$  is not a number, but an element of the set  $S$ . The association rule of a random variable assigns one, and only one real number to each element in the sample space  $S$ . And the range of the random variable is the sample space of numbers defined by the association rule; in short, a set of numbers. The following definition of random variable is given by Papoulis [3].

**Definition 7.** *Random variable  $\chi$  is the function whose domain is the space  $S$ , (i.e., the process of assigning a real number  $\chi(\zeta)$  to each outcome  $\zeta$  of the experiment  $F$ ) such that the set  $\{\chi \leq x\}$  is an event for every real number  $x$  and the probability of the events  $\{\chi = +\infty\}$  and  $\{\chi = -\infty\}$  equals zero, i.e.,  $P[\chi = +\infty] = P[\chi = -\infty] = 0$ .*

It is said that a random variable is *discrete* if the set of possible values of  $\chi$  is countable (though not necessarily finite). Since we do not know for sure the value  $\chi$  will have, we say that it will take the value  $x$  with probability  $p_\chi(x)$ . That is  $p_\chi(x) = P[\chi = x]$ , what is known as *probability mass function* of  $\chi$ . Since a discrete probability distribution is a probabilistic function, it follows the rules and properties of the probability theory. This means that, among other things,  $0 \leq p_\chi(x) \leq 1 \forall x$  and  $\sum_x p_\chi(x) = 1$ .

If we want to know the probability with which  $\chi$  takes the values  $x_i$  less than or equal to any value  $a$ , the following concept is very useful.

**Definition 8.** *The cumulative distribution function of a random variable  $\chi$  is the function  $F_\chi(a) = P[\chi \leq a] = \sum_{x \leq a} p_\chi(x)$ , defined for any real variable  $a$ .*

A random variable is *continuous* if there is a nonnegative function  $f_\chi(x)$ , so the cumulative probability distribution  $F_\chi(x)$  can be calculated as  $F_\chi(a) = P[\chi \leq a] = \int_{-\infty}^a f_\chi(x)dx$ . The function  $f_\chi(x)$  is the *probability density function* of the random variable  $\chi$ . The density function  $f_\chi(x)$  is not necessarily a continuous function but the cumulative distribution function  $F_\chi(x)$  automatically is, implying that  $P[\chi = x] = 0$ , for any value of  $x$ . That means that the events  $\{a \leq \chi < b\}$ ,  $\{a < \chi \leq b\}$  and  $\{a < \chi < b\}$  have the same probability. It is obvious that we can compute the density function from the cumulative distribution function from  $f_\chi(x) = \frac{d}{dx}F_\chi(x)$ .

**Definition 9.** Stochastic process is the family of random variables  $\{\chi_t, t \in T\}$ , where  $T$  is the time parameter space and where for each  $t \in T$ ,  $\chi_t$  is a point in space  $S$ , called state space.

The classification of the stochastic processes, also known as random processes, depends upon the *state space*  $S$  and the nature of the *time parameter space*  $T$ . Both spaces,  $S$  and  $T$ , can be discrete or continuous. If  $T$  is discrete we obtain *stochastic sequences*, but if  $T$  is continuous we have stochastic processes. In this occasion we will only refer to the case of the stochastic sequences known as Markovian chains, which state space is discrete.

### 3 Markov Chains

The Russian mathematician Andrei Andreyevich Markov (1856-1922) is well remembered for his studies on random variables sequences in which the future variable is determined by the present variable, i.e., it only depends upon the previous one. That is why these random variables sequences are called *Markov chains*. The Markov chains are important mainly for two reasons: (i) a wide variety of physical, biological, economic, and social phenomena that can be described through these chains, and (ii) the development of the theory allows us to carry out the calculations required for studying these phenomena.

**Definition 10.** The stochastic sequence  $\{\chi_n | n \in \mathbb{N}\}$ , i.e., the sequence with a discrete parameter space of time  $T$  and a finite or infinite countable state space  $S$  is a Markov chain if  $P[\chi_{n+1} = j | \chi_0 = i_0, \chi_1 = i_1, \dots, \chi_n = i] = P[\chi_{n+1} = j | \chi_n = i]$ .

When the space of states  $S$  is finite, the Markov chain is finite. The expression  $p_{ij} = P[\chi_{n+1} = j | \chi_n = i]$  is the *one-step transition probability* denoting the

probability that the chain goes from state  $i$  to the state  $j$  in just one step<sup>1</sup>. The general form of the  $m$ -step transition probability is  $p_{ij}^m = P[\chi_{n+m} = j | \chi_n = i]$ . It is very important to stress that the  $m$ -step transition probabilities refer to all possible paths in order to get from one state to another. Furthermore, it must be noted that the probabilities  $p_{ij}^m$  must satisfy:

$$0 < p_{ij}^m \leq 1, \quad i, j = 1, 2, \dots, N; \quad m = 0, 1, 2, \dots$$

$$\sum_{j \in S} p_{ij}^m = 1, \quad i = 1, 2, \dots, N; \quad m = 0, 1, 2, \dots$$

**Definition 11.** It is said that a Markov chain is homogeneous in time if it satisfies the condition  $P[\chi_{n+1} = j | \chi_n = i] = P[\chi_{m+1} = j | \chi_m = i] = p_{ij}$ ,  $m \neq n$  is met, i.e., if all transition probabilities  $p_{ij}$  are fixed and do not depend on the current time.

Throughout this paper we will only consider the homogeneous Markov chains.

The probability of going from state  $i$  to state  $j$  in a given number of steps, say  $n + m$ , is the probability of going from state  $i$  to an intermediate state  $k$  in  $n$  steps, and then going to state  $j$  in  $m$  steps. Considering all possible intermediate states  $k$  we obtain  $p_{ij}^{n+m} = \sum_{k \in S} p_{ik}^n p_{kj}^m$ , known as the *Chapman-Kolmogorov equation* for discrete time Markov chains. The matrix form the one-step transition probabilities is:

$$\chi_n \begin{matrix} & \chi_{n+1} \\ & 0 & 1 & 2 & \dots \\ \begin{pmatrix} p_{00} & p_{01} & p_{02} & \dots \\ p_{10} & p_{11} & p_{12} & \dots \\ p_{20} & p_{21} & p_{22} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} & = & \mathbf{P} \end{matrix}$$

Now,  $\mathbf{P}^m = \mathbf{P}^{m-1}\mathbf{P}$ , where  $\mathbf{P}^0 = \mathbf{I}$  is the identity matrix, allows us to calculate the  $m$ -step transition probabilities from the one-step transition matrix powers.

Graph theory is a very useful for studying the Markov chains behavior in an easier way. States are usually denoted by nodes, and arcs show all the possible transitions between states with their respective probabilities. The resulting diagram is called *state transition diagram*.

---

<sup>1</sup> $p_{ij} = p_{ij}^1$ .



But, what happens when the initial state is random? The *state probability*  $\pi_j^m$  of finding the Markov chain in state  $j$  at the  $m$ -th step:  $\pi_j^{(m)} = P[\chi_m = j]$ . We know that  $P[\chi_m = j | \chi_0 = i] = p_{ij}^m$  where we have assumed, without losing of generality, that we are in state  $i$  at time 0. Multiplying both sides of the equation by  $\pi_i^{(0)} = P[\chi_0 = i]$  and summing over all states and applying theorem of total probability, we obtain  $\pi_j^{(m)} = \sum_i \pi_i^{(0)} p_{ij}^m$ , which allows us to compute the distribution of  $\chi_m$  in terms of the initial distribution  $\pi_i^{(0)}$  and the  $m$ -steps transition probabilities.

The row vector formed by the state probabilities at time  $m$  is the probability vector  $\Pi^{(m)} = (\pi_0^{(m)}, \pi_1^{(m)}, \pi_2^{(m)}, \dots)$ . With this definition, the matrix form is  $\Pi^{(m)} = \Pi^{(0)} \mathbf{P}^m$ ,  $m = 1, 2, \dots$

In order to forecast the long-term Markov chain behavior, i.e., when the number of transitions tends to infinity, it is required to study the characteristics of the states that form the chain. That is why we will now consider a classification of Markov states with the following definitions.

Consider states  $i, j \in S$ . If there is a path from  $i$  to  $j$ , that is, if there is an integer  $n$  such that  $p_{ij}^n > 0$ , then we have  $i \rightarrow j$ . It is said that two states *communicate* if there is a path from state  $i$  to state  $j$  and vice versa, denoted by  $i \rightleftharpoons j$ . Let  $C[i] = \{j | i \rightleftharpoons j; j \in S\} \forall i \in S$ . We call  $C[i]$  the *class* of state  $i$ .

**Definition 12.** *The set of states  $C$  of a Markov chain is said to be irreducible if all states in  $C$  communicate with each other state, i.e., if  $p_{ij}^m > 0 \forall m \geq 1, i, j \in C$ .*

If a Markov chain is irreducible it will only have one class of states, i.e.  $C[i] = C[j] \forall i, j \in S$ . Let  $C$  denote any class of state and  $\bar{C}$  be the set of Markov states that are not in class  $C$ .

**Definition 13.** *A class  $C$  is closed if the transition probabilities  $p_{ij}^m = 0 \forall m \geq 1$  where  $i \in S$  and  $j \notin S$ , i.e., if there is not a path from  $C$  to  $\bar{C}$ .*

If  $C$  consists of a single state, say  $i$ , then  $i$  is called an *absorbing state*. A necessary and sufficient condition for  $i$  to be an absorbing state is that  $p_{ii} = 1$ , and obviously  $p_{ij} = 0 \forall i \neq j$ , that is, an absorbing state does not communicate with any other state. On the other hand, all states of an irreducible chain form a closed set and no subset can be closed.

**Definition 14.** A class  $C$  is said to be transient if there is a path out of  $C$ . That is, if  $\exists i \in C$  and  $k \in \bar{C}$  such that  $p_{ik} > 0$ . The states in a transient class are said to be transient.

**Definition 15.** A Markov chain is said to be absorbing if every state in it is either absorbing or transient.

**Definition 16.** A class  $C$  is ergodic if every path that starts in  $C$  remains in  $C$ . So  $\sum_{j \in C} p_{ij} = 1, \forall i \in C$ .

The states in an ergodic class are called *ergodic*. An irreducible Markov chain consists in a particular ergodic class, i.e.,  $C[i] = S, \forall i \in S$ .

**Definition 17.** Let  $T_{ij} = \min\{n \geq 1 : \chi_n = j | \chi_0 = i\}$ <sup>2</sup> be the random variable denoting the time the chain spends to reach, for the first time, to state  $j$  from state  $i$ . The random variable  $T_{ij}$  is called reaching time<sup>3</sup>.

Usually, the definition of the random variable  $T_{ij}$  is not very clear due to the notation used in several books. Definition 17 allows us to understand the meaning of this important concept in an easy way which implies an additional achievement in this work.

The probability of a Markov chain leaving a state  $i$  and *first reaching* state  $j$  in  $n$ -steps is denoted by  $f_{ij}^{(n)} = P\{\chi_n = j, \chi_r \neq j \forall r < n | \chi_0 = i\} = P_i\{T_j = n\}$  and must satisfy  $\sum_{n=1}^{\infty} f_{ij}^{(n)} \leq 1$ . Therefore, the probability of *ever reaching* state  $j$  from state  $i$  is written as<sup>4</sup>:  $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$ . If  $f_{ij} < 1$ , the possibility of state  $i$  *never reaching* state  $j$  exists. If  $f_{ij} = 1$ , state  $i$  will definitely reach *sometime* state  $j$ . Note that when  $\sum_{n=1}^{\infty} f_{ij}^{(n)} = 1$  the random variable  $T_{ij}$  has a probability distribution.

If we want to know the average number of steps required to reach state  $j$  from state  $i$  we must calculate the mean of the distribution  $f_{ij}$ . Let  $\mu_{ij}$  denote the *mean reaching time* from state  $i$  to state  $j$  defined as follows

$$\mu_{ij} = \begin{cases} \infty & \text{if } \sum_{n=1}^{\infty} f_{ij}^{(n)} < 1 \\ \sum_{n=1}^{\infty} n f_{ij}^{(n)} & \text{if } \sum_{n=1}^{\infty} f_{ij}^{(n)} = 1 \end{cases}$$

<sup>2</sup>Note that being in state  $j$  at time 0 is not considered.

<sup>3</sup>When  $i = j$  some authors define  $T_j = \min\{n \geq 1 : \chi_n = j\}$ , as *returning time*.

<sup>4</sup> $f_{ij} \neq f_{ij}^{(1)}$ .

When  $i = j$ , it is said that  $\mu_{ii}$  is the *mean return time*.

**Definition 18.** The period  $d(i)$  of a state  $i$  is the largest number that divides the  $n$  for which  $p_{ii}^n > 0$ , i.e.,  $d(i) = \text{m.c.d.}\{n \geq 1 : p_{ii}^n > 0\}$

**Definition 19.** A state  $i$  is said to be aperiodic if  $d(i) = 1$  and periodic if  $d(i) > 1$ <sup>5</sup>.

Based on the previous definitions, we can now classify the states of a Markov chain as follows:

1. A state is *transient* if  $f_{ii} < 1$ , i.e., if  $\mu_{ii} = \infty$ .
2. A state is *recurrent* if  $f_{ii} = 1$ .
3. A recurrent state is *null* if  $\mu_{ii} = \infty$  and *nonnull* if  $\mu_{ii} < \infty$ <sup>6</sup>.
4. A recurrent state is *ergodic* if it is nonnull and aperiodic.

**Theorem 1.** All states of an irreducible Markov chain have the same type since they can only be either

- all transient
- all recurrent nonnull, or
- all recurrent null.

Furthermore, if states are periodic, then all of them have the same period  $n$ .

The most interesting case for evaluating the behavior of a Markov chain is the one in which the probability distribution  $\pi_j^{(m)}$  does not change when  $m \rightarrow \infty$ . In other words, if after a large number of transitions, regardless the initial state probability  $\pi_j^{(0)}$ , there is a limit probability of finding the chain in state  $i$ .

**Definition 20.** The steady state probability distribution  $\{\pi_j; j \in S\}$  of a Markov chain is given by  $\pi_j = \lim_{m \rightarrow \infty} \pi_j^{(m)}$ .

<sup>5</sup>If  $p_{ii}^n = 0 \forall n \geq 1$ , we say that  $d(i) = 0$ .

<sup>6</sup>An irreducible Markov chain can only have recurrent nonnull states if the number of states is finite.

When this limit does exist, state  $j$  is said to be *stationary* or has reached a *steady state distribution*. The steady state probability does not imply in any way that the chain remains in a particular state, since this one continues having transitions from one state to another with  $p_{ij}^n$  probabilities.

**Theorem 2.** *In an homogeneous, irreducible and aperiodic Markov chain, the limiting probabilities  $\pi_j$  always exist and are independent of the initial state probability distribution  $\pi_j^{(0)}$ . Furthermore, either*

1. *all states are transient or all states are recurrent nonnull. In both cases  $\pi_j = 0 \forall j$  and there is no steady state distribution, or*
2. *all states are recurrent nonnull and then  $\pi_j > 0 \forall j$ , in which case set  $\{\pi_j\}$  is the steady state probability distribution and  $\pi_j = \frac{1}{\mu_j}$ .*

$\pi_j$  probabilities satisfy uniquely:  $\sum_j \pi_j = 1$  and  $\pi_j = \sum_i \pi_i p_{ij}$ .

The equation for  $\pi_j$  can be expressed in matrix form. For this, we define the probability vector  $\Pi$  as  $\Pi = [\pi_0, \pi_1, \pi_2, \dots]$ , and then we can write the equation  $\Pi = \Pi P$ , followed directly from the equation  $\Pi^{(m)} = \Pi^{(m-1)} P$  at the limit when  $m \rightarrow \infty$ .

## 4 Conclusions

Stochastic processes, specifically Markovian chains, have an important application mainly in actuarial sciences, mathematics and engineering. Their study is extensive and requires a wide variety of mathematical theories. It is important to stress that this paper states a logic order of definitions in order to understand the concept of probability under a Boolean algebra approach. Usually, the notation used in several books is difficult for people to read, causing confusion. This article states some basic definitions with a notation that makes possible an easy understanding of the meaning of the probabilistic and stochastic concepts, thus implying an additional achievement.

This work is a brief abstract of my thesis on Actuary degree: “Stochastic Processes and Stochastic Petri Nets” in process. Its purpose is to create the material that provides basic concepts on Markovian Processes and Stochastic Petri nets as an introduction to the system mathematical modeling theory and to spread Petri Nets as a forefront investigation line through an introduction to the system creation, formal analysis and graphic representation study modeled under a random approach.

## 5 Acknowledgement

I would like to thank professors Soledad Carreto, Guillermo A. Vázquez and Zbigniew Oziewicz for inviting me to participate in “The First International Conference on Nonlinear, Fractional and Stochastic Dynamics”, held on January 2005. Moreover, I would like to thank my thesis adviser, Dr. Oziewicz, for all his support. I would also like to thank my friend Martha Mauri for her supervision while translating this work to english.

## References

- [1] Durrett Rick. *Essentials of Stochastic Processes*. Springer-Verlag, New York, 1999.
- [2] Lipschutz Seymour. *Probabilidad*. Mc Graw Hill, México, 1991.
- [3] Papoulis Athanasios. *Probability, Random Variables and Stochastic Processes*. McGraw Hill, New York, 1965.
- [4] Parzen Emanuel. *Stochastic Processes*. Holden Day, San Francisco, 1962.
- [5] Ross Sheldon M. *Stochastic Processes*. John Wiley & Sons, New York, 1996.
- [6] Taylor Howard M. & Karlin Samuel. *An Introduction to Stochastic Modeling*. Academic Press, San Diego, California, 1998.
- [7] Vélez Ibarrola Ricardo. *Procesos Estocásticos*. Universidad Nacional de Educación a Distancia, Madrid, 1996.

# Bibliografía

- [1] Apostol Tom M. *Calculus, Cálculo con funciones de varias variables y álgebra lineal con aplicaciones a las ecuaciones diferenciales y a las probabilidades*. 2a. ed., México, Reverté, 2001.
- [2] Ayres Frank Jr. *Álgebra moderna*. México, McGraw Hill, 1991.
- [3] Bause Falko & Kritzinger Pieter S. *Stochastic Petri Nets, An introduction to the theory*. Wiesbaden, Vieweg & Sohn, 1995.
- [4] Bernstein Stephen & Ruth. *Elements of Statistics I: Descriptive Statistics and Probability*. U.S.A., McGraw Hill, 1994.
- [5] Bronson Richard. *Investigación de operaciones*. México, McGraw Hill, 1993.
- [6] Cameron Peter J. *Sets, Logic and Categories*. Great Britain, Springer, 1999.
- [7] Cases Muñoz Rafael & Márquez Villodre Lluís. *Lenguajes, gramáticas y autómatas*. México, Alfaomega, 2002.
- [8] Cardoso Janette & Camargo Heloisa (Eds.). *Fuzziness in Petri nets*. Heidelberg, Physica-Verlag, 1999.
- [9] Çinlar Erhan. *Introduction to Stochastic Processes*. New Jersey, Prentice Hall, 1997.
- [10] Coleman Rodney. *Procesos estocásticos*. México, Limusa, 1976.
- [11] Devore Jay L. *Probabilidad y estadística para ingeniería y ciencias*. 4a. ed., México, International Thomson Editores, 1998.

- [12] *Diccionario de Términos Científicos y Técnicos*. Barcelona, Mc Graw Hill, 1981.
- [13] Durrett Rick. *Essentials of Stochastic Processes*. New York, Springer-Verlag, 1999.
- [14] Godoy Escoto José Eduardo. *Cuadernos de Investigación, La caracterización de las funciones de probabilidad*. México, Escuela Nacional de Estudios Profesionales Acatlán, 1990.
- [15] González Videgaray MariCarmen. *Modelos y Simulación*. México, Escuela Nacional de Estudios Profesionales Acatlán, 1996.
- [16] González Videgaray MariCarmen. *Números aleatorios, extrapolación y fractales*. México, Centro de Cómputo de la Escuela Nacional de Estudios Profesionales Acatlán, 1995.
- [17] Grimaldi Ralph P. *Discrete and Combinatorial Mathematics, An Applied Introduction*. 3rd. ed., U.S.A., Addison Wesley, 1994.
- [18] Hillier Frederick S. y Lieberman Gerald J. *Investigación de operaciones*. México, McGraw Hill, 2002.
- [19] Johnsonbaugh Richard. *Matemáticas discretas*. 4a. ed., México, Prentice Hall, 1999.
- [20] Lecture Notes in Computer Science. *Application and Theory of Petri Nets 2000*. Denmark, Springer, 2000.
- [21] Lindemann Christoph. *Performance Modelling with deterministic and Stochastic Petri Nets*. England, John Wiley & Sons, 1998.
- [22] Lipschutz Seymour. *Probabilidad*. México, Mc Graw Hill, 1991.
- [23] Mendenhall William, et. al. *Estadística Matemática con Aplicaciones*. 2a. ed., México, Grupo Editorial Iberoamérica, 1994.
- [24] Morales Anel, Navarro Jesús & Pérez Vladimir. *LENA: Un Lenguaje para Especificar Redes de Petri Estocásticas*. México, Escuela Superior de Cómputo, Instituto Politécnico Nacional, 1999.

- [25] Newman James R. *El mundo de las matemáticas*. Barcelona, Grijalbo, 1968.
- [26] Papoulis Athanasios. *Probability, Random Variables and Stochastic Processes*. New York, McGraw Hill, 1965.
- [27] Parzen Emanuel. *Stochastic Processes*. San Francisco, Holden Day, 1962.
- [28] Pérez Ordóñez Edgar, et. al. *Enciclopedia matemática MEGA*. Santafé de Bogotá, Terranova Editores, Tomo III, 1995.
- [29] Real Academia Española. *Diccionario de la Lengua Española*. Madrid, Espasa Calpe, 2001.
- [30] Ross Sheldon M. *Stochastic Processes*. New York, John Wiley & Sons, 1996.
- [31] Sowa John F. *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. U.S.A., Brooks/Cole, 2000.
- [32] Steen S. *La enseñanza agradable de las matemáticas*. México, Noriega-Limusa, Colección de Textos Politécnicos, 1998.
- [33] Stewart Ian. *Does God play dice? The mathematics of chaos*. U.S.A., Blackwell, 1989.
- [34] Taha Hamdy A. *Investigación de operaciones, una introducción*. México, Prentice Hall, 1998.
- [35] Taylor Howard M. & Karlin Samuel. *An Introduction to Stochastic Modeling*. San Diego, California, Academic Press, 1998.
- [36] Tucker Howard G. *Introducción a la teoría matemática de las probabilidades y la estadística*. Barcelona, Vicens-Vives, 1973.
- [37] Vélez Ibarrola Ricardo. *Procesos Estocásticos*. Madrid, Universidad Nacional de Educación a Distancia, 1996.
- [38] Zimmermann Armin, Freiheit Jörn & Hommel Günter. *Discrete Time Stochastic Petri Nets for Modeling and Evaluation of Real-Time Systems*. Berlin, Real-Time Systems and Robotics Group, Technische Universität Berlin, 2000.



## Tesis de Licenciatura sobre procesos estocásticos

1. Herrera Maldonado Mahil. *Introducción a los procesos estocásticos*. Licenciatura en Actuaría, Acatlán, Estado de México, Escuela Nacional de Estudios Profesionales Acatlán, Universidad Nacional Autónoma de México, 2003.
2. Spencer Restrepo Harvey. *Modelo para la descripción de la variación de la abundancia y mortalidad de larvas de peces a través de los procesos estocásticos*. Licenciatura en Actuaría, Acatlán, Estado de México, Escuela Nacional de Estudios Profesionales Acatlán, Universidad Nacional Autónoma de México, 2000.
3. González Videgaray MariCarmen. *Procesos estocásticos y aplicaciones*. Licenciatura en Actuaría, Acatlán, Estado de México, Escuela Nacional de Estudios Profesionales Acatlán, Universidad Nacional Autónoma de México, 1985.

# Índice alfabético

- álgebra
  - booleana de conjuntos, 11
  - de Boole, 10
  - de variables aleatorias, 23
- árbol, 82
  - del conjunto alcanzable, 82
- aditividad finita, 11
- aditividad numerable, 14, 19
- aleatoriedad, 8
- bucle, 78
- cadena, 32
  - de Ehrenfest, 52
  - de Markov, 36
    - absorbente, 44
    - estacionaria, 51
    - finita, 36
    - homogénea, 37
    - irreducible, 43
    - matriz fundamental de, 57
- cardinalidad, 10
- clase de estados, 43
  - absorbente, 44
  - cerrada, 43
  - ergódica, 44
  - transitoria, 44
- coeficiente de variación, 30
- comunicación de estados, 43
- conjunto
  - alcanzable, 81, 82
  - de Borel, 21
  - de eventos exhaustivos, 16
  - potencia, 11
- contacto, 81
- delta de Kronecker, 57
- dependencia estadística, 32
- desviación estándar, 30
- determinismo, 7
- diagrama de transición de estados, 39
- distribución conjunta
  - continua, 31
  - discreta, 30
- distribución de estado estable, 50, 66
- ecuación
  - de Chapman-Kolmogorov, 38
  - diferencial hacia adelante, 65
  - diferencial hacia atrás, 65
  - p-invariante, 88
- espacio
  - de estados, 32
  - muestral, 18
    - finito, 19
    - no numerable, 19
    - numerable, 19
  - parametral de tiempo, 32
- estado
  - absorbente, 43
  - aperiódico, 47

- ergódico, 44, 48
- periódico, 47
- recurrente, 48
  - nulo, 48
  - no nulo, 48
- transitorio, 48
- eventos, 10
  - disjuntos, 16
  - mutuamente excluyentes, 16
  - exhaustivos, 16
- estimación bayesiana, 18
- función
  - de conjunto, 11
    - de aditividad finita, 11
    - de aditividad numerable, 19
  - de distribución acumulativa, 28
  - densidad de probabilidad, 29
    - conjunta, 31
  - masa de probabilidad, 28
- gráfica
  - bipartita, 71
  - dirigida, 71
- grafo del conjunto alcanzable, 83
- hojas, 82
- independencia estadística, 16
- indicador, 25
- intensidad, 63
- marcado, 77
  - acotado, 80
  - hogar, 81
  - seguro, 80, 81
  - vivo, 80
- matriz
  - de incidencia, 86
    - de incidencia hacia adelante, 86
    - de incidencia hacia atrás, 86
    - de tasas de transición, 64
    - de transición, 38
    - doblemente estocástica, 38
    - estocástica, 38
    - fundamental, 57
- medida, 13
  - de probabilidad, 21
    - con aditividad finita, 13
- modelo matemático, 8
- momento
  - primer, 30
  - n-ésimo, 30
  - n-ésimo central, 30
  - segundo central, 30
- número esperado de elementos, 98
- p-invariante, 83
- periodo, 47
- probabilidad, 14
  - abstracta, 27
  - con puntos, 27
  - condicional, 16
  - de alcance, 47
  - de descarga, 98
  - de estado, 41
  - de primera pasada, 45
  - de transición en un paso, 36, 38
  - de transición en m pasos, 36
  - puntual, 20, 27
  - sin puntos, 27
- proceso, 7, 31
  - de Markov, 61
    - homogéneo, 61
    - estocástico, 32
- red de Petri, 71

- acotada, 81
- débilmente conectada, 76
- ecuación fundamental de, 86
- estocástica, 96
  - de tiempo continuo, 96
- Evento-Condición, 81
- fuertemente conectada, 76
- Lugar-Transición, 71
- marcada, 72
- ordinaria, 71
- temporal, 96
- segura, 81
- viva, 81
- relación de flujo, 75
- retículo, 10
  - acotado, 10
  - complementario, 10
  - distributivo, 10
- secuencia de descargas, 77
- subconjuntos medibles, 21
- sucesión de aplicaciones medibles, 32
- suceso, 18
  - elemental, 18
- supervivencia, 80
- t-invariante, 91
- tasa de transición, 63
- transición
  - activada, 77
  - descarga de, 77
  - viva, 81
- teorema
  - de Bayes, 18
  - de probabilidad total, 17
- teoría estocástica, 8
- tiempo
  - de alcance, 45
  - de primera pasada, 45
  - esperado de alcance, 47
  - esperado de recurrencia, 47
- valor esperado, 29
- valor promedio, 29
- variable aleatoria, 24
  - discreta, 27
  - continua, 28
- varianza, 30
- vector de descargas, 87