



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

EL USO DEL ACOPLAMIENTO PARA ESTIMAR LA TASA DE CONVERGENCIA PARA EL ESTADO ESTACIONARIO DE CADENAS DE MARKOV.

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I A

P R E S E N T A :

MARIA JACKELINE CORTES CABRERA



DIRECTOR DE TESIS: DRA. ELIANE REGINA RODRIGUES CALONI

MEXICO, D. F.

DIVISION DE ESTUDIOS PROFESIONALES



2000

FACULTAD DE CIENCIAS SECCION ESCOLAR

277857



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



VENEZUELA NACIONAL
AVENIDA DE
MÉZIZ

MAT. MARGARITA ELVIRA CHÁVEZ CANO

Jefa de la División de Estudios Profesionales

Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis:

El uso del acoplamiento para estimar la tasa de convergencia para el estado estacionario de cadenas de Markov.

realizado por **María Jackeline Cortés Cabrera**

Con número de cuenta **9231828-1**, pasante de la carrera de **Actuaría**

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de tesis

Propietario **Dra. Eliane Regina Rodrigues Caloni** *Eliane Regina Rodrigues*

Propietario **Dr. Eduardo Gutiérrez Peña** *[Signature]*

Propietario **Dr. Mogens Bladt Petersen** *Mogens Bladt*

Suplente **M. en A.P. María del Pilar Alonso Reyes** *[Signature]*

Suplente **Act. Yazmin Iliana Barcenás Orozco** *[Signature]*



Consejo Departamental de **Matemáticas**
FACULTAD DE CIENCIAS
M. en A.P. María del Pilar Alonso Reyes
Coordinadora de Actuaría

“...Que todo lo hagan con amor...”

Co 16, 14

*“Pon tu alegría en el Señor, Él hará lo que desea tu corazón.
Pon tu porvenir en manos del Señor, confía en Él y déjalo actuar”*

Salmo 37 (36), 4-5

Agradecimientos

Este trabajo es dedicado a todas las personas que de algún modo participaron en el logro de esta meta, con su ejemplo, su apoyo, su cariño, sus conocimientos, sus consejos ó con el simple hecho de reconocer que esto es importante para mí y respetarlo, a todos ellos de verdad gracias.

Un especial agradecimiento a mi asesora la Dra. Eliane que no sólo me enseñó a adorar los procesos estocásticos, sino también tener pasión por las cosas que haga y ayudar sinceramente a quien así lo desea.

Es importante señalar también al Dr. Carlos Castillo Chávez de quien tome ejemplo de persistencia. Al Dr. Manuel Mendoza y al Dr. Gorostiza de quienes recibí palabras de aliento y ejemplo.

A mis sinodales, Yazmin, Pilar, Eduardo, Mogens, gracias por toda su paciencia y sus aportaciones.

A mis profesores que durante la carrera sembraron mi gusto por la probabilidad y la estadística.

Agradezco muy especialmente a mi familia por todo su amor y a la memoria de los que ya no están conmigo.

No debo dejar de mencionar a mis queridos amigos Roberto, Christian, Roxana, Isabel, Karina, Mabel, Ricardo, Jessica, Angy (del CIMAT); Vanessa; al grupo de Biomatemáticas: Era, Judy, Leo, Chucho... A mis compañeros en el IMATE: Diana, Mito, Gerardo, Daniel, Rogelio, Cruz, Mayam, Mika, Juan Carlos.... A mis compañeros y amigos de la generación de Actuaría: Alberto, Aline, Alvaro, Bety, Eliel, Gerardo, Hugo, Jana, Juan Carlos, Laura, Malú, Marbella, Nadia, Pablo,... A Adolfo por su cariño y apoyo.

Un reconocimiento a la gente que trabaja en el IMATE y en la Fundación ADA por el apoyo que ofrecen a los becarios.

Índice General

Introducción	v
1 Introducción a Cadenas de Markov	1
1.1 Introducción	1
1.1.1 Cadenas de Markov	5
1.2 Cadenas de Markov a tiempo discreto	6
1.2.1 Clasificación de estados	12
1.3 Cadenas a Tiempo Reversible	25
1.4 Cadenas de Markov a tiempo continuo	27
1.4.1 Ecuaciones diferenciales de Kolmogorov.	31
1.4.2 Probabilidades límite para una cadena de Markov con- tinua	38
2 Acoplamiento y cadenas de Markov	45
2.1 Introducción	45
2.2 Distancia entre distribuciones	46
2.2.1 Distancia de variación total	46
2.2.2 Distancia de separación total	49
2.3 Tiempo aleatorio y tiempo de paro	52
2.4 Método de acoplamiento	53
2.5 Convergencia exponencial	61
3 Un problema de estructura de datos	65
3.1 Introducción	65
3.2 Movimiento al frente	66
3.3 Barajeos	68
3.4 Distribuciones estacionarias	71
3.5 Acoplamiento	76

3.6	El coleccionador de cupones	78
3.7	Resultados Principales	80
4	Comentarios	89
4.1	Introducción	89
4.2	Probabilidades no uniformes	89
4.3	Sucesiones de demandas	91
4.4	Tiempo conocido y tiempo analizable	92
4.5	Otros barajeos	92
4.6	Monte Carlo vía cadenas de Markov	93
5	Apéndice	95
5.1	Desigualdad de Chebyshev	95
5.2	Distribución geométrica	97
	Bibliografía	101

Introducción

La contemplación de la naturaleza, el anhelo de poseerla, los rumbos que han tomado las necesidades del hombre, la satisfacción de estas necesidades y el placer de tener respuestas, fueron motivos para crear a las diversas ciencias. Las matemáticas son fruto de estas inquietudes, ellas resuelven y respaldan con bastante efectividad ciertos cuestionamientos. La herramienta de la cual se sirven las matemáticas traduce hechos de interés a cierto lenguaje, dónde estos hechos son reinterpretedados bajo estos nuevos términos. Así, el modelo matemático de un fenómeno es una descripción, para su implementación necesita al fenómeno, los cuestionamientos a su respecto, un sistema bajo el cual analizar y conexiones lógicas entre el hecho real y el implementado de forma artificial.

A los modelos matemáticos se les ha clasificado en deterministas o estocásticos. Independientemente de ello, al modelo se le juzga por su utilidad y ésta, a su vez, es determinada por la facilidad con la que responde a la necesidad planteada. Así, una misma situación puede tener dos o más formas de plantearse dependiendo de la agudeza del observador en su búsqueda de detalles útiles y de su creatividad.

Un modelo estocástico predice un conjunto de posibles respuestas, basándose en las probabilidades de ciertos aspectos involucrados en la formulación del sistema artificial.

El objetivo general de este trabajo es encontrar el tiempo requerido (exacto o aproximado), en el cual un proceso llega y permanece con ciertas características conocidas y/o deseables. La búsqueda de este tiempo se ve motivada por lo práctico que resultaría el realizar cualquier proceso sin tenerlo que vigilar exhaustivamente, sino sólo a partir del tiempo en que se incluye

la característica deseada; además, existe mayor interés si se agrega que no dependerá este tiempo de condiciones iniciales.

En concreto la incógnita a resolver de esta tesis es: encontrar la n^* , a partir de la cual una cadena de Markov se encuentre lo más cerca posible del estado estacionario (estado de equilibrio); o dicho de otro modo que la distribución estacionaria y la distribución de la cadena de Markov al tiempo n^* guarden una distancia (distancia de variación total) lo más insignificante posible. En muchos casos esta distancia no puede ser medida; cuando así sucede, el acoplamiento puede jugar un papel muy importante, pues ésta es una técnica que permite hacer la comparación entre medidas de probabilidad construyendo en un mismo espacio muestral lo necesario, donde sus distribuciones de probabilidad son las medidas de probabilidad a comparar. La tarea del acoplamiento consiste en definir una cota superior para la diferencia existente entre la distribución al tiempo n^* y la distribución estacionaria.

La probabilidad y la estadística han utilizado este método como herramienta en diversos estudios de donde se han creado diferentes variantes del método de acuerdo a su fin, entre los que se encuentran: acoplamiento de Ornstein, acoplamiento Mineka, acoplamiento simple, acoplamiento débil, acoplamiento de procesos regenerativos, entre otros. (Ver Lindvall(1992) para mayores detalles). El método de acoplamiento se inició en 1938 con la publicación del trabajo de Doeblin, desde entonces varios autores han contribuido con sus aportaciones; de entre ellos mencionamos Griffeath (1975), Goldstein (1979), Aldous (1983) y a Lindvall (1992). En éste último se cuenta con una recopilación de la técnica de acoplamiento, se recomienda esta referencia como una fuente de resultados y trabajos hasta hoy expuestos sobre la técnica de acoplamiento.

El ejemplo que se planteará como aplicación del acoplamiento considera un arreglo con $N > 1$ registros, donde cada uno de ellos puede ser solicitado independientemente de los demás, con cierta distribución de probabilidad. Después de satisfacer la solicitud, este registro es regresado y colocado al inicio del arreglo, es decir, en la primera posición de la izquierda. Esta forma de reacomodo define varias posibilidades en las que se encuentra el arreglo. Y la pregunta a contestar es: ¿en qué tiempo dos arreglos que actúan bajo el mismo esquema (distribuidos como ya se mencionó) tendrán el mismo orden en sus registros?

Como se pudo apreciar, las herramientas clave para este trabajo son las siguientes: cadenas de Markov, acoplamiento, distancia de variación y la forma de solución del problema del coleccionador de cupones; en conjunto darán lugar a resultados muy interesantes.

- El primer capítulo introduce el entorno de las cadenas de Markov. Inicia desde la definición de un proceso estocástico hasta la correspondiente a cadena de Markov continua, pasando por la de cadena de Markov discreta, cadena de Markov reversible y algunas particularidades útiles de las mismas. Es decir, este capítulo es de conceptos preliminares.
- El segundo capítulo expone la técnica de acoplamiento para cadenas de Markov y los conceptos de distancia y tiempo de convergencia para cadenas de Markov, agregando un breve apartado de convergencia exponencial.
- El tercer capítulo muestra una aplicación de lo estudiado en los anteriores capítulos a un esquema de estructura de datos con ciertas particularidades utilizando a el problema del coleccionador de cupones para averiguar la n^* a partir de la cual la cadena de Markov diseñada para esta aplicación se encuentra en el estado estacionario. Finalmente, se muestran resultados de lo más interesante y concluyente para todo el trabajo.
- El capítulo cuatro es compuesto por comentarios en relación a varios puntos a lo largo de la tesis.
- El capítulo cinco informa sobre resultados de probabilidad aplicados en ciertas secciones del trabajo.

Capítulo 1

Introducción a Cadenas de Markov

“Hay series ideales de acontecimientos que corren paralelos a los reales. Rara vez coinciden, por lo general, los hombres y las circunstancias modifican la serie ideal perfecta y sus consecuencias son por lo tanto igualmente imperfectas . . . ”

Novalis.

1.1 Introducción

Este capítulo tiene a bien ser tomado como una introducción al tema de esta tesis: las cadenas de Markov.

La estructura de este primer capítulo llevará definiciones básicas en probabilidad, conceptos, condiciones y teoremas de cadenas de Markov que aseguran la convergencia de éstas al estado estacionario. Todo ello con el fin de establecer una buena exposición de capítulos posteriores, planteando de esta forma una línea de trabajo que va de lo general a lo particular. Es importante mencionar que se considera sólo los temas que guardan una relación directa con el tópico que da título a este material. Si se desea profundizar en el tema se sugiere en primera instancia la consulta de la bibliografía.

Definición 1.1.1. *Cualquier hecho creado y observado cuidadosamente de tal forma que puede ser recreado bajo las mismas condiciones es llamado*

experimento; cuando este experimento es reproducido en varias ocasiones y obtiene diversas respuestas es del tipo **aleatorio**.

Definición 1.1.2. El conjunto de todos los posibles resultados de un experimento es llamado **espacio muestral**. El espacio muestral será denotado por el símbolo Ω .

Por ejemplo,

$$\Omega = \{\omega : \omega = \text{diferentes cartas de un mazo de 52}\},$$

de otro modo,

$$\Omega = \{\omega : \omega = x_i, i = 1, 2, 3 \dots 52\}, \text{ donde } x_i = i\text{-ésima carta.}$$

La realización del experimento tiene varias posibles respuestas que son medidas por el siguiente concepto.

Definición 1.1.3. Sea Ω espacio muestral y $A_i \subset \Omega$. Se dice que P es una **medida de probabilidad sobre Ω** , cuando es un operador que mapea una colección de elementos de Ω a $[0, 1]$, satisfaciendo

- $P(A_i) \geq 0$ para todo $A_i \subset \Omega$.
- $P(\emptyset) = 0$.
- Para A_1, A_2, \dots subconjuntos de Ω tales que $A_i \cap A_j = \emptyset$, para toda $i \neq j$ se tiene que $P(\bigcup_i A_i) = \sum_i P(A_i)$.

Definición 1.1.4. Una clase $\mathfrak{F} \neq \emptyset$ de subconjuntos de Ω es una **σ -álgebra sobre Ω** si

1. $\Omega \in \mathfrak{F}$.
2. Si $A \in \mathfrak{F}$, entonces $A^c \in \mathfrak{F}$.
3. Para $A_1, A_2, \dots \in \mathfrak{F}$ existen entonces $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{F}$.

Observación: Una medida de probabilidad sobre \mathfrak{F} hace el mapeo de \mathfrak{F} a $[0, 1]$, tal que $P(\Omega) = 1$. Además de que $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$, para una secuencia $A_n \in \mathfrak{F}, n \geq 1$, donde los elementos son ajenos.

Definición 1.1.5. Si Ω es un espacio muestral finito o infinito numerable, \mathfrak{F} es la σ -álgebra sobre \mathfrak{F} y P es una medida de probabilidad sobre Ω , entonces $(\Omega, \mathfrak{F}, P)$ es llamado **espacio de probabilidad**.

Definición 1.1.6. Para Ω espacio muestral y \mathfrak{F} una σ -álgebra de Ω , entonces (Ω, \mathfrak{F}) es llamado un **espacio medible**.

Definición 1.1.7. Sea $(\Omega, \mathfrak{F}, P)$ espacio de probabilidad, $\Omega = \{\omega_1, \omega_2, \dots, \omega_k\}$, \mathfrak{F} σ -álgebra sobre Ω , P medida de probabilidad sobre \mathfrak{F} y $A \subset \Omega$. La **probabilidad del evento** A , $P(A)$ es definida como

$$P(A) = \sum_{i=1}^k P(\omega_i).$$

En particular $P(\omega_i) = p_i$. El conjunto vacío es definido con probabilidad cero.

Definición 1.1.8. Para $\Omega = \mathbb{R}$, la σ -álgebra de **Borel sobre \mathbb{R}** , indicada por $B(\mathbb{R})$ es la menor σ -álgebra generada por todos los conjuntos abiertos de \mathbb{R} .

Definición 1.1.9. Sean Ω y Ω' dos espacios muestrales y \mathfrak{F} y \mathfrak{F}' las respectivas σ -álgebras, $f : \Omega \rightarrow \Omega'$, para cada $A \subset \Omega'$. Sea $f^{-1}(A) = \{\omega \in \Omega : f(\omega) \in A\}$ la inversa de f . Entonces una transformación medible de (Ω, \mathfrak{F}) a (Ω', \mathfrak{F}') si $f^{-1}(\mathfrak{F}') \subset \mathfrak{F}$. Si $(\Omega', \mathfrak{F}') = (\mathbb{R}, B(\mathbb{R}))$, entonces f es llamada **simplemente medible en (Ω, \mathfrak{F})** .

El espacio de probabilidad (Ω, P) contiene todo lo necesario para el análisis del experimento. Para tender un lazo en el lenguaje entre quienes tratan con mayor familiaridad con el ambiente del experimento y los que no, además de sumar a ello un mayor énfasis en el estudio de características específicas de los resultados, se realiza un mapeo de Ω a los reales, definiendo:

Definición 1.1.10. Sea (Ω, \mathfrak{F}) y $(\mathbb{R}, B(\mathbb{R}))$ dos espacios medibles. Una **variable aleatoria** es un mapeo $X : \Omega \rightarrow \mathbb{R}$, tal que para alguna $A \in B(\mathbb{R})$, $X^{-1}(A) \in \mathfrak{F}$.

Si X es una variable aleatoria sobre $(\Omega, \mathfrak{F}, P)$. La medida de probabilidad de X , P_X sobre $B(\mathbb{R})$ es

$$P_X(A) = P\{\omega | X(\omega) \in A\}.$$

Definición 1.1.11. La función de distribución de una variable aleatoria X es la función $F = F_X$ que va de los \mathbb{R} a $[0, 1]$ dada por

$$F(X) = P\{\omega | X(\omega) \leq x\}, x \in \mathbb{R}$$

Observación: Estos conceptos son elementales en la teoría de la probabilidad pueden ser encontrados, por ejemplo, en Bosq (1996) y Ash (1972).

Definición 1.1.12. Un **proceso estocástico** es una familia de variables aleatorias $X = \{X_t, t \in T\}$, donde T es el conjunto índice y $X_t \in S$ para todo $t \in T$, donde S es el espacio de estados ó espacio donde la variable aleatoria X_t es definida.

Se puede clasificar a los procesos estocásticos de acuerdo a las características de T y S . Estos son discretos o continuos.

Si el conjunto de índices T es numerable infinito, entonces $X = \{X_t : t \in T\}$ indica un proceso estocástico con parámetro discreto ó sucesión estocástica ó proceso estocástico a tiempo discreto. Si T es un intervalo, acotado o no, $X = \{X_t : t \in T\}$ indica un proceso estocástico con parámetro continuo ó proceso estocástico a tiempo continuo. El espacio de estados S puede ser infinito (numerable o no) o finito. Su expresión tal vez sea unidimensional, bidimensional ó n-dimensional. Debido a la clasificación de los anteriores conjuntos se tiene que los procesos estocásticos pueden ser:

	Espacio de estados		
Tiempo	finito	infinito numerable	infinito no numerable
discreto	x	x	x
continuo	x	x	x

Algunos tipos usuales de procesos estocásticos importantes de mencionar son:

1. Cadenas (Procesos) de Markov: Son sucesiones (procesos) estocásticas, donde dado el presente, su comportamiento futuro y pasado son independientes.
2. Procesos Bernoulli: $X = \{X_n : n = 1, 2, \dots\}$ con $S = \{0, 1\}$, X_n , $n = 1, 2, \dots$ variables aleatorias independientes y con la misma distribución

$$P(X_n = 1) = p = 1 - P(X_n = 0), \quad \text{para toda } n = 1, 2, \dots$$

3. Procesos con incrementos independientes: Es un proceso estocástico $X = \{X_t : t \in T\}$, tal que, para toda n y para todo $t_1 < t_2 < t_3 < \dots < t_n \in T$, $X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ son independientes.
4. Procesos con incrementos estacionarios: Es un proceso estocástico $X = \{X_t, t \in T\}$, tal que para $t, s \in T$ con $t < s$, la distribución de $X_s - X_t$ depende sólo de $s - t$, esto es que la distribución de los cambios entre los valores de dos variables depende de la distancia que guardan entre ellas.
5. Procesos de Conteo: Es un proceso $X = \{X_t, t \in T\}$ que representa el número de ocurrencias de eventos al tiempo t y satisface $X_t \geq 0$, $X_t \in \mathbb{Z}$, si $s < t$ entonces $X_s < X_t$. Para $s < t$, $X_s - X_t$ es el número de eventos ocurridos en $(s, t]$.
6. Martingalas: Es un proceso $\{X_n : n \geq 1\}$, tal que $E[|X_n|] < \infty$ para toda n y $E[X_{n+1} | X_1, \dots, X_n] = X_n$, donde $E(X)$ representa la esperanza de X y $E(X|Y)$ es la esperanza condicional de X dado Y , para X, Y variables aleatorias en un mismo espacio de estados.

1.1.1 Cadenas de Markov

Las siguientes secciones introducirán al lector a la teoría básica de las cadenas de Markov. Se espera que con estos conceptos se otorguen algunas de las herramientas necesarias para abordar el problema de estimación del tiempo de convergencia para el estado estacionario de una cadena de Markov.

Primeramente nuestra atención se fijará en las cadenas de Markov a tiempo discreto, posteriormente serán presentados algunos resultados respecto al caso continuo.

A continuación se definirá formalmente lo que es un proceso y una cadena de Markov, donde T es discreto o continuo.

Definición 1.1.13. *El proceso estocástico $X = \{X_t, t \in T\}$ con espacio de estados S es llamado **proceso de Markov**, si cumple la propiedad de Markov, es decir, para algunas $t, s \in T$ con $t < s$, la distribución condicional de X_s dado X_t , es la misma que la distribución condicional de X_t dado $\{X_u, u \leq t\}$. Formalmente, para $x_u, u \leq t$ y $A \subset S$,*

$$P(X_s \in A | X_u = x_u, u \leq t) = P(X_s \in A | X_t = x_t) \quad (1.1)$$

Cuando S es finito o infinito numerable, el proceso es llamado **cadena de Markov**.

Definición 1.1.14. Una **cadena de Markov de orden r** es aquel proceso estocástico $X = \{X_t, t \in T\}$ con espacio de estados S , finito o numerable, donde el valor de la cadena en el tiempo t depende de exactamente de los r estados anteriores, es decir, para $t_1 < t_2 < \dots < t_n \in T$, $x_1, \dots, x_n \in S$, $A \subset S$ y $r < n$,

$$\begin{aligned} P(X_{t_n} \in A | X_{t_1} = i_1, \dots, X_{t_{n-1}} = i_{n-1}) & \quad (1.2) \\ & = P(X_{t_n} \in A | X_{t_{n-1}} = i_{n-1}, \dots, X_{t_{n-r}} = i_{n-r}) \end{aligned}$$

1.2 Cadenas de Markov a tiempo discreto

Sean algunos ejemplos:

1. Considere un sistema en el cual dos piezas de equipo idénticas son instaladas en paralelo. Estas piezas de equipo actúan independientemente una de la otra y en un día tienen una disponibilidad $\alpha \in [0, 1]$. De inicio se encuentran en buenas condiciones. Sea el interés en el número de piezas de equipo que falla durante n días, siendo más específicos, sea $X_n, n \geq 0$, el número de piezas de equipo que no está en buenas condiciones al principio del n -ésimo día, de donde $X_0 = 0, 1, 2$.

Este proceso tiene un espacio de estados $S = \{0, 1, 2\}$ finito y T discreto.

En este ejemplo, se ha observado que la predicción del estado de la pieza en el tiempo futuro, depende solamente del estado de la pieza actualmente. Cualquier dato del pasado no aporta información adicional.

2. Sea supuesto un sitio de taxis, donde los clientes llegan a formarse a una fila para solicitar el servicio y abordar la unidad. Cada taxi llega de manera independiente de acuerdo con un intervalo de tiempo exponencialmente distribuido. Cada cliente es atendido por período. Sea supuesto que los clientes arriban a la fila independientemente y de acuerdo a un intervalo de tiempo exponencialmente distribuido. Sea X_n el número de clientes esperando en fila al inicio del período n .

Este proceso $X_n = \{X_n, n \geq 0\}$ tiene un espacio de estados numerable y usando Y_n para denotar el número de nuevos clientes arribando durante el n -ésimo período, se tiene:

$$X_{n+1} = \begin{cases} X_n - 1 + Y_n, & \text{si } X_n \geq 1 \\ Y_n, & \text{si } X_n = 0 \end{cases}$$

Al igual que en el caso anterior, para determinar el número de clientes esperando en la fila se debe saber cuántos han llegado justo en el momento anterior y/o cuántos se tenían esperando en la fila. De manera formal, se dice que dado Y_n sucesión de variables aleatorias independientes e idénticamente distribuidas, cuando se buscan las probabilidades de transición a un paso se formula lo siguiente:

$$\begin{aligned} P(X_{n+1} = j | X_n = 0) &= P(Y_n = j), \\ P(X_{n+1} = j | X_n = i \neq 0) &= P(Y_n = j - i + 1) \end{aligned}$$

para toda $n \geq 0$. Por lo que es un claro ejemplo de cadena de Markov a tiempo discreto.

Definición 1.2.1. Una cadena de Markov $X = \{X_n, n \geq 0\}$ con espacio de estados S (finito o numerable) es llamada **cadena de Markov a tiempo discreto**, si el conjunto de tiempo T es discreto, es decir, $T = \{0, 1, 2, \dots\}$, para toda $n \geq 0$ e $i_1, \dots, i_n \in S$.

Observaciones:

A lo largo de esta tesis...

1. se trabajará con espacios de estados finitos o numerables y con cadenas de Markov de primer orden, a menos que se mencione algún otro tratamiento.
2. al leer el término "estados" se hace alusión a las posibles respuestas de los eventos que puede tomar cada fase de nuestro proceso.
3. la **distribución inicial** de una cadena de Markov discreta $X = \{X_n : n = 0, 1, \dots\}$ con espacio de estados S será notada de la siguiente forma

$$P(X_0 = j) = \pi_0(j). \quad (1.3)$$

Una probabilidad de transición es la probabilidad de trasladarse de un estado a otro en un único paso, pudiendo conservar cierta dependencia con respecto al tiempo donde ocurre el cambio.

Definición 1.2.2. Sea $X = \{X_n : n = 0, 1, \dots\}$ una cadena de Markov con espacio de estados S . Sea definido la **probabilidad de transición a un paso** como

$$P_{ij}^{(n, n+1)} = P(X_{n+1} = j \mid X_n = i)$$

leyendo esto como la probabilidad de estar en el estado j en el tiempo $n + 1$ dado que en el tiempo n se está en el estado i , para todo $i, j \in S$.

Definición 1.2.3. Sea $X = \{X_n : n = 0, 1, 2, \dots\}$ cadena de Markov, con espacio de estados S , definimos a la **probabilidad de transición a m pasos**,

$$P_{ij}^{(n, n+m)} = P(X_{n+m} = j \mid X_n = i).$$

Esto es, la probabilidad de estar en ese estado j al tiempo $n + m$ dado que en el tiempo n se está en el estado i .

Definición 1.2.4. La cadena de Markov $X = \{X_n : n = 0, 1, \dots\}$ con espacio de estados S , es una **cadena de Markov homogénea en el tiempo**; si tiene como propiedad característica que su probabilidad de transición sea independiente del tiempo, dicho de otro modo, donde ocurre el cambio de estado sólo depende del número de pasos necesario para ir de un estado al otro, es decir,

$$P_{ij}^{(n, n+m)} = P(X_{n+m} = j \mid X_n = i) = P_{ij}^{(m)}$$

para toda $i, j \in S$. En el caso de $m = 1$ se tiene que $P_{ij}^{(n, n+1)}$ es indicado por P_{ij} .

Cuando se trabaja con cadenas de Markov a tiempo discreto, es decir, cuando T es discreto se puede entender su movimiento en el tiempo como el número de pasos que se realizan. Así entonces se estaría de acuerdo al decir que si se trabaja con una cadena de Markov homogénea, nuestras probabilidades de transición dependen sólo del número de pasos y no del tiempo del cual parten.

Un arreglo matricial puede contener información que caracteriza la naturaleza de una cadena de Markov de forma que se tuviera un resumen de ella.

Definición 1.2.5. *Será llamada **matriz de transición** ó **matriz de probabilidades de transición** al arreglo matricial que contenga las probabilidades de transición como elementos. Así, se tendrá que de la cadena de Markov $X = \{X_n : n = 0, 1, \dots\}$ con espacio de estados S , la matriz de transición a un paso será*

$$P = (P_{ij}^{(1)})_{i,j \in S}.$$

Y la matriz de transición de n pasos será

$$P^{(n)} = (P_{ij}^{(n)})_{i,j \in S},$$

note que,

$$0 \leq P_{ij}^{(n)} \leq 1$$

y que

$$\sum_{j \in S} P_{ij}^{(n)} = 1.$$

Observación: Una matriz de transición a un paso con espacio de estados $S = \{0, 1, 2, \dots, i\}$ es de la siguiente forma:

$$\begin{pmatrix} P_{00} & P_{01} & P_{02} & P_{03} & \dots & P_{0i} \\ P_{10} & P_{11} & P_{12} & P_{13} & \dots & P_{1i} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{i0} & P_{i1} & P_{i2} & P_{i3} & \dots & P_{ii} \end{pmatrix}$$

Tal que, P_{00} es la probabilidad de transición a un paso, pasando del estado 0 al estado 0, siguiendo así, la P_{i3} es la probabilidad de transición a un paso, pasando del estado i al estado 3 y así se sigue.

Dada la naturaleza de las cadenas de Markov es posible obtener una fórmula recursiva para las probabilidades de transición en $m \geq 2$ pasos. Esto se obtiene utilizando el siguiente teorema:

Teorema 1.2.6. Ecuaciones de Chapman-Kolmogorov

Sea P la matriz de transición de $\{X_n : n = 0, 1, 2, \dots\}$, una cadena de Markov homogénea con espacio de estados S , entonces:

$$P_{ij}^{(n)} = \sum_{k \in S} P_{ik}^{(r)} P_{kj}^{(n-r)}$$

para $0 \leq r \leq n$, $n \geq 1$ y

$$P_{ij}^{(0)} = \begin{cases} 1, & \text{si } i = j \\ 0, & \text{si } i \neq j. \end{cases}$$

Demostración. Sea $r \geq 0$ y $n \geq 1$

$$\begin{aligned} P_{ij}^{(n)} &= P(X_n = j \mid X_0 = i) = \frac{P(X_n = j, X_0 = i)}{P(X_0 = i)} & (1.4) \\ &\stackrel{(1)}{=} \sum_{k \in S} \frac{P(X_n = j, X_r = k, X_0 = i)}{P(X_0 = i)} \\ &\stackrel{(2)}{=} \sum_{k \in S} P(X_n = j \mid X_r = k, X_0 = i) \frac{P(X_r = k, X_0 = i)}{P(X_0 = i)} \\ &\stackrel{(2)}{=} \sum_{k \in S} P(X_n = j \mid X_r = k, X_0 = i) P(X_r = k \mid X_0 = i) \\ &\stackrel{(3)}{=} \sum_{k \in S} P(X_n = j \mid X_r = k) P(X_r = k \mid X_0 = i) \\ &\stackrel{(4)}{=} \sum_{k \in S} P_{ik}^{(r)} P_{kj}^{(n-r)} \end{aligned}$$

- (1). Ley de probabilidad total.
- (2). Def. de probabilidad condicional.
- (3). Propiedad de Markov.
- (4). Def. de probabilidad de transición. □

Observación: Por el teorema 1.2.6 y por las propiedades de multiplicación entre matrices se puede escribir, para S finito,

$$P^{(n)} = P^n$$

donde P^n es la n -ésima potencia de P .

La demostración de este hecho es por inducción. Sea indicado por P_{ij}^n el

ij -ésimo elemento de la matriz P^n . De este modo para $n = 2$, por el teorema 1.2.6 se tiene que, para S finito.

$$P_{ij}^{(2)} = \sum_{k \in S} P_{ik} P_{kj} \stackrel{(1)}{=} P_{ij}^2, \text{ entonces } P^{(2)} = P^2.$$

(1). Por definición de multiplicación de matrices.

Suponga válido para $n = k$, es decir,

$$P_{ij}^{(k)} = P^k$$

y se demostrará que la igualdad es válida para $n = k + 1$. De este modo

$$P_{ij}^{(k+1)} \stackrel{(1)}{=} \sum_{k \in S} P_{ik} P_{kj}^{(k)} \stackrel{(2)}{=} \sum_{k \in S} P_{ik} P_{kj}^k = P_{ij}^{k+1}$$

entonces $P^{(k+1)} = P^{k+1}$ y la igualdad vale para todo $n \geq 2$. Para $n = 1$, la igualdad es trivial.

(1). Por el teorema 1.2.6.

(2). Definición de multiplicación de matrices.

Notas:

1. Se hace la diferencia en la notación P^z y $P^{(z)}$, donde la primera es la matriz $\underbrace{P \cdots P}_z$, mientras que la segunda es una matriz de transición a z pasos.

2. Se puede calcular $\pi_n(\cdot) = P(X_n = \cdot)$ usando las siguientes fórmulas recursivas.

Lema 1.2.7. Sea $\pi_0(\cdot) = P(X_0 = \cdot)$ y $\pi_n(\cdot) = P(X_n = \cdot)$. Entonces

$$(i). \pi_n(k) = \sum_{j \in S} \pi_0(j) P_{jk}^{(n)}$$

o equivalentemente

$$(ii). \pi_n(k) = \sum_{j \in S} \pi_{n-1}(j) P_{jk}$$

Demostración. (i).

$$\begin{aligned}\pi_n(k) = P(X_n = k) &= \sum_{j \in S} P(X_n = k, X_0 = j) & (1.5) \\ &= \sum_{j \in S} P(X_n = k | X_0 = j) \pi_0(j) \\ &= \sum_{j \in S} \pi_0(j) P_{jk}^{(n)}\end{aligned}$$

(ii).

$$\begin{aligned}\pi_n(k) = P(X_n = k) &= \sum_{j \in S} P(X_n = k, X_{n-1} = j) & (1.6) \\ &= \sum_{j \in S} P(X_n = k | X_{n-1} = j) \pi_{n-1}(j) \\ &= \sum_{j \in S} \pi_{n-1}(j) P_{jk}\end{aligned}$$

□

1.2.1 Clasificación de estados

Me era imposible explicar la razón de esa diferencia, salvo que supusiera que los objetos destrozados eran los que habían sido completamente absorbidos, mientras que los otros habían penetrado en el remolino en un período más adelantado de la marea, o bien, para luego ser absorbidos, que no habían alcanzado a tocar el fondo del vórtice antes del cambio del flujo y del reflujó, según fuera el momento.

Edgar Allan Poe, El descenso al Maelstrom.

Hasta ahora ya se tiene una breve semblanza de la probabilidad de que la cadena se encuentre en un estado dado un cierto número finito de pasos. Pero ¿qué pasará, cuándo $n \rightarrow \infty$?, ¿cómo será esta probabilidad de transición?, ¿importará de dónde parte?, ¿cuál es el comportamiento de cada estado? Los siguientes conceptos ayudarán a responder estas preguntas.

Sea $X = \{X_n : n = 0, 1, \dots\}$ una cadena de Markov e $i, j \in S$ dos estados cualesquiera.

Definición 1.2.8. Un estado j es **accesible** desde el estado i , si para alguna $n \geq 0$, $P_{ij}^{(n)} > 0$.

Es decir, un estado j es accesible desde el estado i , cuando partiendo del estado i puede llegar al estado j , después de un número n de pasos.

Definición 1.2.9. Se dice que **dos estados i y j se comunican**, cuando i es accesible desde j y viceversa. La notación a usar es $i \longleftrightarrow j$ (i se comunica con j). Equivalentemente existen $n, m \geq 0$, tal que, $P_{ij}^{(n)} > 0$ y $P_{ji}^{(m)} > 0$.

Proposición 1.2.10. El concepto de comunicación es una relación de equivalencia. Es decir, para $i, j, k \in S$, valen las siguientes propiedades.

1. Reflexividad. $i \longleftrightarrow i$.
2. Simetría. Si $i \longleftrightarrow j$, entonces $j \longleftrightarrow i$.
3. Transitividad. Si $i \longleftrightarrow j$ y $j \longleftrightarrow k$, entonces $i \longleftrightarrow k$.

Demostración. 1. $i \longleftrightarrow i$, note que dado que la cadena esta en estado i , entonces para $n = 0$ se tiene $P_{ii}^{(0)} = 1$, entonces $i \longleftrightarrow i$.

2. Por definición de comunicación, que dado $i \longleftrightarrow j$, tal que, $P_{ij}^{(n)} > 0$, $P_{ji}^{(m)} > 0$ existen $n, m \geq 0$ y por tanto $j \longleftrightarrow i$.

3. Si $i \longleftrightarrow j$ entonces por definición de comunicación existen $n, m \geq 0$, tal que, $P_{ij}^{(n)} > 0$, $P_{ji}^{(m)} > 0$ y si $j \longleftrightarrow k$, entonces existen $n_1, m_1 \geq 0$, tal que, $P_{kj}^{(n_1)} > 0$, $P_{jk}^{(m_1)} > 0$, se desean $n^*, m^* \geq 0$, tal que, $P_{ki}^{(m^*)} > 0$ y $P_{ik}^{(n^*)} > 0$. Sea $n^* = n + m_1$, entonces por las ecuaciones de Chapman-Kolmogorov se tiene que,

$$P_{ik}^{(n^*)} = \sum_{l \in S} P_{il}^{(n)} P_{lk}^{(m_1)}.$$

Por otro lado, $P_{ij}^{(n)} P_{jk}^{(m_1)} > 0$, entonces

$$P_{ik}^{(n^*)} = \sum_{l \in S} P_{il}^{(n)} P_{lk}^{(m_1)} \geq P_{ij}^{(n)} P_{jk}^{(m_1)} > 0$$

para ciertas $n, m_1 \geq 0$. De manera análoga se demuestra que $P_{ki}^{(m^*)} > 0$ para $m^* = m + n_1$.

□

Al existir estas relaciones de equivalencia, se tendrán clases de equivalencia, donde los estados en una clase de equivalencia se comunican entre sí. Un espacio de estados puede tener una o más clases de equivalencia dependiendo de cómo los estados se relacionan entre sí.

Definición 1.2.11. *Se dice que una cadena es irreducible, si la relación de equivalencia dada por la proposición 1.2.10 produce una y sólo una clase de equivalencia.*

Para la clasificación de los estados de una cadena de Markov se deben saber algunos datos a cerca de ellos. Se preguntarán cosas como por ejemplo, ¿cuántas veces la cadena de Markov regresa a un estado específico?... por lo que se introduce en nuestro tema el concepto de período.

Definición 1.2.12. *Sea P una matriz de transición de una cadena de Markov $X = \{X_n : n = 0, 1, \dots\}$ con espacio de estados S . El estado $i \in S$ es llamado **periódico de período $d(i)$** , si*

$$d(i) = \text{mcd}\{n : P_{ii}^{(n)} > 0\}$$

Si $d(i) = 1$, entonces es llamado **aperiódico**. Es decir, $d(i)$ es el valor tal que $P_{ii}^{(n)} = 0$ para toda $n \neq d(i), 2d(i), \dots$ y $d(i)$ es el mayor número con esta propiedad. Si $P_{ii}^{(n)} = 0$ para toda $n \geq 1$, entonces no hay regreso a ese estado.

Nota: mcd indica máximo común divisor.

Se expondrán posteriormente los conceptos de irreductibilidad y aperioidicidad, los cuales tendrán un lugar de importancia para el estudio de convergencia al estado de equilibrio para una cadena de Markov X .

Uniendo las propiedades de las clases de equivalencia y las de periodicidad, se tienen los siguientes resultados.

Proposición 1.2.13. *Sean $i, j \in S$, dos estados de una cadena de Markov descrita X . Si $i \longleftrightarrow j$, implica $d(i) = d(j)$. Es decir, el período es una propiedad de clase.*

Demostración. Dado que por hipótesis $i \longleftrightarrow j$ entonces por definición, existen $n, m \geq 0$, tal que $P_{ij}^{(n)} > 0$ y $P_{ji}^{(m)} > 0$, por lo que $P_{ij}^{(n)} P_{ji}^{(m)} > 0$. Ahora, tome $N = n + m$; entonces por el teorema 1.2.6

$$P_{ii}^{(N)} \geq P_{ij}^{(n)} P_{ji}^{(m)} > 0.$$

Sea una $w \geq 1$, tal que $P_{jj}^{(w)} > 0$, entonces de las ecuaciones de Chapman-Kolmogorov, se tiene que $P_{ii}^{(N+w)} \geq P_{ij}^{(n)} P_{jj}^{(w)} P_{ji}^{(m)} > 0$.

Sea $d(i)$, el período de estado i , entonces por la definición de período y dado que $P_{ii}^{(N)} > 0$ y $P_{ii}^{(N+w)} > 0$, se tiene que $d(i)$ divide N y $N + w$. Por esta razón divide a todo w tal que $P_{jj}^{(w)} > 0$. Por tanto, $d(i)$ divide a $d(j)$. Usando un argumento análogo para el período de j , se tiene que $d(j)$ divide $d(i)$. De este modo $d(i) = d(j)$. \square

Con el período se sabe en cuántos pasos la cadena tiene probabilidad positiva de regresar al estado de dónde partió, también se puede saber cuándo regresa, por ejemplo, cuando lo hace por primera vez.

A seguir, se tendrán dos definiciones que auxiliarán a identificar cuándo una cadena de Markov es recurrente o transiente.

Definición 1.2.14. Sea $X = \{X_n : n = 0, 1, \dots\}$ una cadena de Markov con espacio de estados S . Sea considerado al estado $i \in S$ fijo, entonces sea definido la probabilidad de regresar por primera vez al estado i en el n -ésimo paso ($n \geq 1$), dado que se partirá del estado i al tiempo inicial por:

$$f_{ii}^{(n)} = P\{X_n = i, X_v \neq i, v = 1, 2, \dots, n-1 \mid X_0 = i\},$$

y la probabilidad de regresar al estado j en el n -ésimo paso dado que se partirá del estado i

$$f_{ij}^{(n)} = P\{X_n = j, X_v \neq j, v = 1, 2, \dots, n-1 \mid X_0 = i\}.$$

En particular, cuando $n = 1$:

$$f_{ii}^{(1)} = P\{X_1 = i \mid X_0 = i\} = P_{ii},$$

$$f_{ij}^{(1)} = P\{X_1 = j \mid X_0 = i\} = P_{ij},$$

mientras que cuando $n = 0$, será definido

$$f_{ij}^{(0)} = \begin{cases} 1, & \text{si } i = j. \\ 0, & \text{si } i \neq j. \end{cases}$$

Definición 1.2.15. La *probabilidad de que partiendo del estado i se llegue al estado j en algún tiempo $n \geq 1$* es definida por,

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

Observación: Note que se puede escribir:

$$P_{ij}^{(n)} = \sum_{k=0}^n f_{ij}^{(k)} P_{jj}^{(n-k)}, n \geq 1$$

para toda $i, j \in S$.

El espacio de estados estará clasificado en diversos subconjuntos; aunque puede ser un único conjunto si la cadena es irreducible. El tipo de la cadena de Markov dependerá de la característica que tienen los estados contenidos en su espacio de estados.

La clasificación se hará de forma variada (en el caso irreducible). Por tanto, se ofrece una clasificación preliminar para los estados de una cadena de Markov discreta. (Se dice preliminar, por que las siguientes características no son las únicas, agregando que el caso próximo (i) puede mostrar subdivisiones, como se estudiará adelante).

Definición 1.2.16. Sea S el espacio de estados de X la cadena de Markov descrita y $j \in S$, entonces

- i. Un estado j es **recurrente**, si $f_{jj} = 1$.
- ii. Un estado j es **transiente**, si $f_{jj} < 1$.

Se entenderá por estado recurrente, áquel del que se está seguro de que al partir de ese estado dado de la cadena, regresará a él mismo en algún paso futuro. Cuando no se tiene esa certeza, el estado será uno del tipo transiente.

Teorema 1.2.17. Sea X una cadena de Markov con S el espacio de estados e $i \in S$. Entonces un estado i es transiente si y sólo si

$$(i). \sum_{n=1}^{\infty} P_{ii}^{(n)} < \infty.$$

Un estado i es recurrente si y sólo si

$$(ii). \sum_{n=1}^{\infty} P_{ii}^{(n)} = \infty.$$

Demostración. Se demostrarán ambas afirmaciones al mismo tiempo. Si el estado i es recurrente, entonces con probabilidad 1 la cadena X retornará al estado i , de donde partirá nuevamente y retornará a i , una y otra vez. De donde, se establece que el número de visitas al estado i es infinito y por tanto que su esperanza también.

El estado i es recurrente si y sólo si,

$$E(\text{número de visitas que } X \text{ hace a } i | X_0 = i) = \infty$$

Sean $i \in S$, un estado transiente e Y la variable aleatoria que registra el número de veces que la cadena X regresa a i antes de dejar de hacerlo dado que partió de i . Por definición (1.2.15) se tiene que la probabilidad de que X regrese a i dado que partió de i es f_{ii} . De este modo se tiene que Y tiene distribución geométrica con probabilidad de éxito $(1 - f_{ii})$.

$$P(Y = n) = f_{ii}^n (1 - f_{ii}), n \geq 0$$

y

$$E[Y] = \frac{1}{1 - f_{ii}}$$

De este modo

$$E[Y] = \frac{1}{1 - f_{ii}} < \infty,$$

si y sólo si $f_{ii} < 1$. Por tanto i es transiente si y sólo si

$$E(\text{número de visitas que } X \text{ hace a } i | X_0 = i) < \infty.$$

Se define la siguiente variable aleatoria indicadora

$$I_n = \begin{cases} 1, & \text{si } X_n = i, \\ 0, & \text{en otro caso,} \end{cases}$$

por lo que $\sum_{n=0}^{\infty} I_n$ es el número de visitas a i y de este modo

$$E\left[\sum_{n=0}^{\infty} I_n | X_0 = i\right] = \sum_{n=0}^{\infty} E[I_n | X_0 = i] = \sum_{n=0}^{\infty} P_{ii}^{(n)},$$

de donde si el estado i es recurrente la suma es ∞ y si i es transiente la suma es finita. \square

Corolario 1.2.18. *Si j es recurrente e $i \longleftrightarrow j$, entonces i es recurrente, es decir, recurrencia es una propiedad de clase.*

Demostración. Como $i \longleftrightarrow j$, entonces por definición de comunicación existen $n, m \geq 0$, tal que $P_{ij}^{(n)} > 0, P_{ji}^{(m)} > 0$. Además si j es recurrente se tiene por el teorema 1.2.17,

$$\sum_{n=1}^{\infty} P_{jj}^{(n)} = \infty.$$

Sea $s \geq 0$, entonces por las ecuaciones de Chapman-Kolmogorov (teorema 1.2.6) y siguiendo la demostración de la proposición 1.2.10,

$$P_{ii}^{(s+m+n)} \geq P_{ij}^{(n)} P_{jj}^{(s)} P_{ji}^{(m)}$$

Sumando sobre s y tomando en cuenta que j es recurrente.

$$\sum_{s=1}^{\infty} P_{ii}^{(s+m+n)} \geq P_{ij}^{(n)} \left(\sum_{s=1}^{\infty} P_{jj}^{(s)}\right) P_{ji}^{(m)},$$

dado que $P_{ij}^{(n)} P_{ji}^{(m)} > 0$ y que j es recurrente se tiene que $\sum P_{ii}^{(s+m+n)} = \infty$ y por tanto i es recurrente. \square

Observación: Una consecuencia del teorema 1.2.17 y del corolario 1.2.18 es que el ser estado transiente también es una propiedad de clase. De este modo, se tiene que los estados que se comunican tienen la misma clasificación.

Definición 1.2.19. *Sea S el espacio de estados de una cadena de Markov X . Sea C una clase de equivalencia en S (inducida por el concepto de comunicación, definición 1.2.9). Se dice que C es una clase cerrada, si es imposible trasladarse afuera de ella, es decir que, $P_{jk} = 0$, para todo $j \in C$ y $k \notin C$.*

Se tiene el siguiente teorema para una clase de estados recurrentes. Un resultado equivalente es válido para la clase de estados transientes.

Proposición 1.2.20. *Sea C una clase de equivalencia en S . Si C es una clase de estados recurrentes, entonces es una clase cerrada.*

Demostración. Sea C una clase de estados recurrentes. Tome $i \in C$ y $j \notin C$. Dado que $j \notin C$, entonces i y j no se comunican. Suponga que existe $k \geq 1$, tal que $P_{ij}^{(k)} > 0$ y dado que $j \notin C$, entonces $P_{ji}^{(n)} = 0$ para todo $n \geq 1$ (caso contrario $i \longleftrightarrow j$ y j estaría en C). De este modo, si se inicia en i jamás se regresa a i . Esto es absurdo dado que i es recurrente. Entonces $P_{ij}^{(k)} = 0$ para todo $k \geq 1$, por lo que C es cerrada. \square

La clasificación de estado recurrente se puede hacer más fina aún. Para mostrar ésto se define:

Definición 1.2.21. *Sea X una cadena de Markov discreta con espacio de estados S . Sea i un estado recurrente. Será llamado **tiempo medio de recurrencia al estado i** a la esperanza del número de pasos a dar hasta el primer retorno a i , es decir, para*

$$R_i = \min\{n \geq 1 | X_n = i\}$$

se tiene $E[R_i | X_0 = i] = \sum_{n=1}^{\infty} n f_{ii}^{(n)}$. El tiempo medio de retorno a i será denotado por μ_i .

Por lo pronto, se verá cómo utilizarlo para la clasificación de los estados de una cadena de Markov discreta.

Una muestra de la utilidad de este concepto la puede dar el teorema siguiente, que se enuncia sin presentar la demostración.

Teorema 1.2.22. *Sea considerada $X = \{X_n, n = 0, 1, \dots\}$ una cadena de Markov recurrente, irreducible y aperiódica. Dado $X_0 = i$, entonces*

$$\lim_{n \rightarrow \infty} P_{ii}^{(n)} = \frac{1}{\sum_{n=0}^{\infty} n f_{ii}^{(n)}} = \frac{1}{\mu_i}.$$

Nota: Este teorema puede ser consultado en libros de procesos estocásticos, se recomiendan en la bibliografía a Karlin, Ross ó Feller en sus respectivos capítulos de cadenas de Markov y/o teoremas límite.

Sean entonces los siguientes criterios de clasificación de estados.

Definición 1.2.23. *Sea X una cadena de Markov con espacio de estados S . Tome $i \in S$ y sea μ_i el tiempo medio de retorno a i , entonces*

- i. el estado i es **recurrente positivo**, si $\mu_i = E[R_i | X_0 = i] < \infty$.
- ii. y el estado i es **nulo recurrente** si $\mu_i = \infty$.

Definición 1.2.24. La **distribución estacionaria de una cadena de Markov** con matriz de transición P y espacio de estados S es una colección de números $\{\pi(i), i \in S\}$ que satisface

$$\begin{aligned} \pi(i) &\geq 0, & \sum_{i=0}^{\infty} \pi(i) &= 1 \\ \pi(j) &= \sum_{i=0}^{\infty} \pi(i) P_{ij}, & j &= 0, 1, 2, \dots \end{aligned}$$

Teorema 1.2.25. Sea C una clase recurrente positiva, aperiódica, con estados $j = 0, 1, 2, \dots$. Entonces existen números $\pi(0), \pi(1), \dots$, estrictamente positivos, tales que,

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{ij}^{(n)} &= \pi(j) \\ \sum_{j \in S} P_{ij} &= 1 \end{aligned}$$

independiente del valor inicial i y $(\pi(1), \dots, \pi(j), \dots)$ es determinada por el conjunto de ecuaciones

$$\begin{aligned} \pi(i) &\geq 0, & \sum_{i=0}^{\infty} \pi(i) &= 1 \\ \pi(j) &= \sum_{i=0}^{\infty} \pi(i) P_{ij}, & j &= 0, 1, 2, \dots \end{aligned}$$

Nota: Este teorema puede ser consultado en libros de procesos estocásticos, se recomiendan en la bibliografía a Karlin, Ross ó Feller en sus respectivos capítulos de cadenas de Markov y/o teoremas límite.

Observaciones:

(i). Note que para una cadena de Markov X con todos los estados recurrentes positivos y aperiódicos, entonces se tiene que X cuenta con una distribución límite, es decir, si $i, j \in S$ espacio de estados y si $P_{ij}^{(n)}$ es la probabilidad en n pasos para X , entonces el límite

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)}$$

existe y es estrictamente positivo y no depende del estado inicial i . Es exactamente en el estudio de este límite que se estará interesado. Es deseable saber cuando se puede parar una cadena de Markov de modo que su distribución es "muy cercana" (a ser especificado posteriormente) de su distribución en equilibrio (o distribución límite).

(ii). Note que si la cadena X es irreducible, entonces C es única y si es una clase recurrente positiva y aperiódica, entonces existe una única distribución límite y esta coincide con la distribución estacionaria de la cadena X .

(iii). Las ecuaciones de la definición 1.2.24 son también conocidas como **ecuaciones de balance total**.

(iv). Saber el tipo de los estados capacitará para determinar si una cadena de Markov converge al equilibrio donde la distribución límite es estrictamente positiva para todo $i \in S$.

Proposición 1.2.26. *Sea $X = \{X_n : n = 0, 1, \dots\}$ una cadena de Markov con espacio de estados S , matriz de transición $P = (P_{ij})_{i,j \in S}$, distribución estacionaria $\pi(\cdot)$ y distribución inicial $\pi_0(\cdot)$. Si la distribución de probabilidad inicial es la distribución estacionaria de X , entonces X_n , tendrá la misma distribución para toda $n \geq 0$*

Demostración. Se utilizará inducción matemática. Sea $j \in S$, entonces para $n = 1$.

$$\begin{aligned}\pi_1(j) &= \sum_{i \in S} P(X_0 = i)P(X_1 = j | X_0 = i) \\ &= \sum_{i \in S} \pi_0(i)P_{ij} \stackrel{(1)}{=} \sum_{i \in S} \pi(i)P_{ij} \stackrel{(2)}{=} \pi(j)\end{aligned}$$

(1). Por hipótesis se tiene $\pi_0(i) = \pi(i)$, para toda $i \in S$.

(2). Definición de distribución estacionaria.

Se ha supuesto que el resultado es válido para $n - 1$, es decir $\pi_{n-1}(i) = \pi(i)$, para todo $i \in S$ y se va a demostrar que el resultado vale para n .

$$\begin{aligned}\pi_n(j) &= \sum_{i \in S} P(X_n = j | X_{n-1} = i)\pi_{n-1}(i) \\ &= \sum_{i \in S} P_{ij}\pi_{n-1}(i) \\ &\stackrel{(1)}{=} \sum_{i \in S} P_{ij}\pi(i) \\ &\stackrel{(2)}{=} \pi(j)\end{aligned}$$

(1). Hipótesis de inducción.

(2). Por definición de distribución estacionaria. (def. 1.2.24)

□

De esta demostración se puede establecer el siguiente corolario,

Corolario 1.2.27. Sea $X = \{X_n : n = 0, 1, \dots\}$ una cadena de Markov con espacio de estados S y distribución estacionaria $\pi(\cdot)$. Si el estado inicial X_0 es seleccionado de acuerdo a la distribución estacionaria, entonces la distribución de probabilidad conjunta (X_{n-1}, X_n) es dada por

$$P(X_n = j, X_{n-1} = i) = \pi(i)P_{ij}$$

Demostración. Tome $i \in S$, por hipótesis del corolario se tiene que

$$\pi_0(i) = P(X_0 = i) = \pi(i)$$

Entonces se tiene que para todo $i, j \in S$

$$\begin{aligned} P(X_n = j, X_{n-1} = i) &\stackrel{(1)}{=} P(X_n = j | X_{n-1} = i)P(X_{n-1} = i) \\ &= P(X_n = j | X_{n-1} = i)\pi_{n-1}(i) \stackrel{(2)}{=} \pi(i)P_{ij} \end{aligned}$$

□

(1). Por condicionalidad.

(2). Por la proposición 1.2.26 y definición de probabilidad de transición 1.2.2.

Se puede ahora empezar a tomar un rumbo más específico en las características de las cadenas de Markov en las cuales se está interesado. De este modo es definido:

Definición 1.2.28. Sea nombrada *cadena de Markov ergódica* a la cadena de Markov X irreducible con espacio de estados S , cuyos estados son recurrentes positivos y aperiódicos.

Observaciones: Es deseable hacer un resumen de algunos puntos que serán de importancia al analizar la convergencia al equilibrio para una cadena de Markov.

1. Una distribución límite es $\pi(j) = \lim_{n \rightarrow \infty} P_{ij}^{(n)}$ y ésta es independiente de la distribución inicial. Nuestro interés será estimar el tiempo n^* de modo que para $n \geq n^*$, $\pi_n(\cdot)$ y $\pi(\cdot)$ son muy próximas (en el sentido considerado en el siguiente capítulo). Note que decir que $\pi_n(\cdot)$ y $\pi(\cdot)$

son muy próximas es lo mismo que decir que $P_{ij}^{(n)}$ es muy próxima de $\pi(\cdot)$, dado que por el lema 1.2.7,

$$\pi_n(j) = \sum_{k \in S} P_{kj}^{(n)} \pi_0(k)$$

se toma el límite de esta expresión y como existe $\lim_{n \rightarrow \infty} P_{ij}^{(n)}$, entonces se tiene que

$$\lim_{n \rightarrow \infty} \pi_n(j) = \sum_{k \in S} \pi(j) \pi_0(k) = \pi(j)$$

2. La distribución estacionaria satisficará el sistema

$$\pi(i) \geq 0, \sum_{i \in S} \pi(i) = 1, \pi(j) = \sum_{i=0}^{\infty} \pi(i) P_{ij},$$

para $j = 0, 1, \dots$ e inversamente cada solución es una distribución estacionaria de la cadena de Markov.

3. Cuando existe una distribución límite, ésta es siempre una distribución estacionaria.
4. Pero, una distribución estacionaria no siempre es una distribución límite. Por ejemplo, para la cadena de Markov periódica siguiente, cuya P es

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

se afirma que

$$P^{(n \text{ par})} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad P^{(n \text{ impar})} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

por lo que el límite de la matriz de transición no existe. Para n par y n impar se puede decir $n = 2m$ y $n = 2m + 1$, $m = 0, 1, \dots$, respectivamente. Se demostrará que $P^{(2m)}$ y $P^{(2m+1)}$ tienen la forma arriba

presentada. La demostración será hecha por inducción. De este modo, se divide en dos casos, para $m = 0, 1, \dots$

(i). $n = 2m$ y (ii). $n = 2m + 1$, $m = 0, 1, \dots$

(i). $n = 2m$.

Para $m = 0$ es trivial dado que por definición

$$P^{(0)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Sea valido para $m = k$, es decir para $n = 2k$. De este modo

$$P^{(2k)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Se demostrará que la expresión es valida para $m = k + 1$, es decir para $n = 2(k + 1)$. Note que,

$$P^{(2(n+1))} = P^{(2n+2)} \stackrel{(1)}{=} P^{(2n)} P^{(2)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

(1). Por el teorema 1.2.6 de las ecuaciones de Chapman-Kolmogorov y la observación correspondiente.

(ii). Para n impar la demostración es similar.

Note que

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{ii}^{(2n)} &= 1 & \text{y} & & \lim_{n \rightarrow \infty} P_{ij}^{(2n)} &= 0, \text{ cuando } j \neq i \\ \lim_{n \rightarrow \infty} P_{ii}^{(2n+1)} &= 0 & \text{y} & & \lim_{n \rightarrow \infty} P_{ij}^{(2n+1)} &= 1, \text{ cuando } j \neq i \end{aligned}$$

De este modo, se tiene que $\lim_{n \rightarrow \infty} P_{ii}^{(2n)} \neq \lim_{n \rightarrow \infty} P_{ii}^{(2n+1)}$ para todo i y para todo $j \neq i$, $\lim_{n \rightarrow \infty} P_{ij}^{(2n)} \neq \lim_{n \rightarrow \infty} P_{ij}^{(2n+1)}$. Por lo que se concluye diciendo que no existe la distribución límite

$$\lim_{n \rightarrow \infty} P_{ij}^{(n)}.$$

Note que $\Pi = (\frac{1}{2}, \frac{1}{2})$ es la distribución estacionaria para la cadena de Markov con matriz de transición P , dado que el sistema de ecuaciones $\pi P = \pi$ es dado por,

$$\pi(2) = \pi(1) \quad (1.7)$$

$$\pi(1) = \pi(2) \quad (1.8)$$

$$\pi(1) + \pi(2) = 1. \quad (1.9)$$

Despejando de la ecuación (1.9), se tiene que $\sum_{i \in S} \pi(i) = 1$

$$\pi(2) = 1 - \pi(1)$$

y sustituyendo en las ecuaciones (1.7) ó (1.8) resulta que

$$\pi(1) = 1 - \pi(1)$$

y por tanto

$$\pi(1) = \frac{1}{2} = \pi(2)$$

por lo cual $\Pi = (\pi(1), \pi(2)) = (\frac{1}{2}, \frac{1}{2})$ cumple con las condiciones de una distribución estacionaria.

5. Una cadena de Markov $X = \{X_n, n \geq 0\}$ se dice **estacionaria**, si las distribuciones de las variables aleatorias X_n son idénticas para todo $n \geq 0$.

1.3 Cadenas a Tiempo Reversible

Considere la siguiente situación: suponga tener un proceso de producción automatizada de cierto objeto; si se dejase a las máquinas trabajando a solas durante cierto período y al tiempo que se retornase, sea observado que los artículos tienen cierto defecto, que se va agravando en cada objeto que va después de este, de tal forma que el último está bastante alejado del prototipo. La primera reacción es averiguar a partir de qué momento empezó todo el problema. La información que tienes es lo que pasa en el presente y lo que te interesa es la probabilidad de ocurrencia de cierto evento pasado.

Las preguntas son entonces: ¿qué puedo saber de mi pasado, a partir de mi presente?, ¿bajo cuáles condiciones?

En los fenómenos de modelado utilizando cadena de Markov a tiempo reversible el interés es el mismo. Algunas veces es más fácil entender lo que pasa en el presente analizando las diversas situaciones del pasado que pudieran dar origen a lo que ocurre, de tal modo que surge la siguiente definición:

Definición 1.3.1. Sea $X = \{X_n, n = 0, 1, 2, \dots\}$ una cadena de Markov estacionaria con espacio de estados S , con $P_{ij}, i, j \in S$ su probabilidad de transición de un paso y $\pi(j), j \in S$, su probabilidad estacionaria. Sea definida la cadena reversa en el tiempo con respecto a P_{ij} como la cadena $X^* = \{X_n^* : n = 0, 1, 2, \dots\}$ con espacio de estados S y matriz de transición $P^* = (P_{ij}^*)_{i, j \in S}$, dada por

$$P_{ij}^* = \frac{\pi(j)P_{ji}}{\pi(i)} \quad (1.10)$$

Observación: La cadena X^* con matriz de transición P^* tiene distribución estacionaria $\{\pi(i), i \in S\}$. Para observar esto note que,

$$\begin{aligned} \sum_{i \in S} \pi(i)P_{ij}^* &= \sum_{i \in S} \pi(i) \frac{\pi(j)P_{ji}}{\pi(i)} \\ &= \sum_{i \in S} \pi(j)P_{ji} \\ &= \pi(j) \sum_{i \in S} P_{ji} \end{aligned}$$

y como $\sum_{i \in S} P_{ji} = 1$

$$\sum_{i \in S} \pi(i)P_{ij}^* = \pi(j).$$

Definición 1.3.2. Cuando $P_{ij} = P_{ij}^*$ para toda $i, j \in S$, la cadena de Markov X se dice que es reversible en el tiempo con respecto a su distribución estacionaria π .

Proposición 1.3.3. *Una cadena de Markov X con espacio de estados S , matriz de transición P y distribución estacionaria π es reversible con respecto a π , si y sólo si,*

$$\pi(j)P_{ji} = \pi(i)P_{ij}.$$

Nota: Este resultado puede ser encontrado en varios libros. Se sugiere a Ross en el capítulo de cadenas de Markov.

Observaciones:

(1). La proposición 1.3.3 dice que una cadena de Markov reversible es tal que el proceso se mueve con la misma velocidad, tanto si va de i a j o de j a i en el tiempo, para todo $i, j \in S$.

Proposición 1.3.4. *Si la cadena de Markov X es reversible con respecto a una distribución ν , entonces ν es la distribución estacionaria de X .*

Demostración. Sea S el espacio de estados de X y suponga que X es reversible con respecto a ν . Entonces por la proposición 1.3.3

$$\nu(j)P_{ji} = \nu(i)P_{ij},$$

Ahora, sumando con respecto a todo $j \in S$ se tiene que,

$$\sum_{j \in S} \nu(j)P_{ji} = \sum_{j \in S} \nu(i)P_{ij} = \nu(i) \sum_{j \in S} P_{ij} = \nu(i)$$

y dado que ν es una distribución, entonces $\sum_{i \in S} \nu(i) = 1$. Entonces por la definición 1.2.24, se tiene que ν es la distribución estacionaria de X . \square

1.4 Cadenas de Markov a tiempo continuo

Entre los diversos procesos estocásticos existe uno cuya particularidad reside en la introducción del concepto de períodos aleatorios de tiempo entre la realización de cambios de estado. Para su definición se necesita de algunos otros conceptos que serán presentados líneas abajo.

Definición 1.4.1. Sea $X = \{X_t, t \geq 0\}$ un proceso estocástico a tiempo continuo, con espacio de estados $S = \{0, 1, 2, \dots\}$. El proceso X es un **proceso Semi-Markov** si cada vez que entra al estado $i \in S$

- el próximo estado será j con probabilidad $P_{ij}, j \in S, \sum_{j \in S} P_{ij} = 1$.
- dado que el próximo estado será $j \in S$, el tiempo gastado en i antes de cambiar a j tiene distribución F_{ij} .

Observación: Se debe hacer notar que en el concepto de proceso Semi-Markov, para hacer cualquier tipo de predicción se tiene necesidad del período de tiempo gastado en cada estado.

Con lo anterior y con ayuda de lo revisado en las cadenas de Markov a tiempo discreto, se abordará este tema con mayor soltura.

Sea un proceso estocástico T , en particular $T = [0, \infty)$ y su conjunto de valores S finito o infinito numerable.

Definición 1.4.2. Será llamada **cadena de Markov a tiempo continuo** al proceso $X = \{X_t : t \in [0, \infty)\}$, tal que para todo $s, t \geq 0$ e $i, j, x_u \in S, 0 \leq u < s$, se tiene

$$P(X_{t+s} = j | X_s = i, X_u = x_u, 0 \leq u < s) = P(X_{t+s} = j | X_s = i) = P_{ij}(s, t),$$

y si

$$P_{ij}(s, t) = P_{ij}(t), \quad i, j \in S, \quad t \geq 0$$

entonces X es **homogénea en el tiempo**.

De este modo es respetada la independendencia entre los estados pasados y los estados futuros, dado los estados presentes.

Una cadena de Markov de tiempo continuo también puede ser definida de la siguiente forma:

Definición 1.4.3. Una **cadena de Markov de tiempo continuo** $X = \{X_t : t \in [0, \infty)\}$ es un proceso semi-Markov, tal que

1. dado que la cadena X entra en el estado $i \in S$, el próximo estado será $j \in S, j \neq i$ con probabilidad P_{ij} , donde $\sum_{j \neq i} P_{ij} = 1$.

2. F_{ij} es una distribución exponencial con parámetro ν_i , es decir para τ_i el tiempo gastado en i antes de pasar a $j \neq i$, se tiene

$$P(\tau_i > t) = \exp(-\nu_i t).$$

Observación: Un estado i tal que $\nu_i = \infty$ es llamado estado instantáneo, pues el abandono del estado es inmediato. Si el estado es tal que $\nu_i = 0$, entonces i es absorbente.

Al introducir una medida de tiempo, se puede saber la rapidez de transición de un estado a otro, que es de lo cual se encargan las tasas. Así pues la tasa de transición es la rapidez con la que pasa de un estado a cualquier otro diferente.

Definición 1.4.4. Sea $X = \{X_t : t \in [0, \infty)\}$ una cadena de Markov continua como en la definición 1.4.3. De este modo para $i, j \in S, i \neq j$ se tiene que la tasa de transición de i a j , es definida por

$$q_{ij} = \nu_i P_{ij}. \quad (1.11)$$

Observación: Note que por la definición de tasa de transición de i a j , $i \neq j$ y la definición de cadena de Markov continua, se tiene que

$$\sum_{j \neq i} q_{ij} = \sum_{j \neq i} \nu_i P_{ij} = \nu_i \sum_{j \neq i} P_{ij} = \nu_i$$

Algunos ejemplos de cadena de Markov continua son los siguientes:

1. Procesos de nacimiento y muerte

Un proceso de nacimiento y muerte es una manera de describir el número de individuos con los que cuenta una población. Este mecanismo será descrito por una cadena de Markov a tiempo continuo con espacio de estados $S = \{0, 1, \dots\}$ para la cual la tasa de transición para $i, j \in S, i \neq j$ será

$$q_{ij} = 0,$$

siempre que $|i - j| > 1$.

Esta última condición permitirá un sólo nacimiento o una sola muerte a cada período de tiempo considerado. Las tasas de transición serán renombradas de la siguiente manera $q_{i,i+1} = \lambda_i$, tasa de nacimiento; $q_{i,i-1} = \mu_i$, tasa de muerte. Cuando $i = 0$,

$$\mu_0 = 0.$$

Para $S' = \{0, 1, \dots, N\}$ se tiene que agregar la condición $\lambda_N = 0$.

Dado que $\nu_i = \sum_{j \neq i} q_{ij}$ y que $q_{ij} = 0$ para $j \neq i-1, i+1$, se obtiene que, $\nu_i = \mu_i + \lambda_i$, con $i \geq 0$.

Para determinar nuestras probabilidades de transición se ha tomado,

$$q_{ij} = P_{ij}\nu_i, \quad \text{para toda } i \neq j.$$

Entonces

$$P_{ij} = \frac{q_{ij}}{\nu_i}.$$

Para este caso, con $i \in S$, se tiene

$$P_{i,i+1} = \frac{q_{i,i+1}}{\nu_i} = \frac{\lambda_i}{\lambda_i + \mu_i}, \quad i = 0, 1, 2, \dots$$

y

$$P_{i,i-1} = \frac{q_{i,i-1}}{\nu_i} = \frac{\mu_i}{\lambda_i + \mu_i} \quad \text{con } i = 0, 1, \dots$$

Observación: Existen...

1. algunas variantes de este proceso, donde la muerte y el nacimiento son actividades que incrementan o decrementan el número de algún objeto de estudio o activan/desactivan cierta actividad.

Sea un sistema $M/M/s$, es decir un sistema donde se actúa según la notación usada: la primera letra M representa el tiempo entre llegadas de clientes que es distribuido exponencialmente, la segunda letra M representa la distribución del servicio dado, que también es exponencial y s es el número de filas del sistema.

Los tiempos entre llegadas son exponencialmente distribuidos con media $\frac{1}{\lambda}$. Cada cliente que llegue al sistema será atendido si se encuentra un servidor disponible; en caso de que no lo haya, se forma una fila de espera. El servicio es sucesivo. Se considerará que el servicio y la llegada de clientes es mutuamente independiente. Los tiempos sucesivos de servicio son variables aleatorias exponenciales con media $\frac{1}{\mu}$. Sea X_t al número de ocupantes del sistema al tiempo t . El proceso $\{X_t, t \geq 0\}$, tendrá la siguiente regla de correspondencia en relación a sus dinámicas

de "nacimiento" (entrada al sistema) y "muerte" (salida al sistema). Sea $\eta_n, n \geq 1$ la tasa de "muerte", si existen n individuos en el sistema. Entonces para

$$\eta_n = \begin{cases} n\mu & \text{si } 1 \leq n \leq s \\ s\mu & \text{si } n > s \end{cases}$$

Sea λ_n la tasa de "nacimiento" si existen n individuos en el sistema. Entonces $\lambda_n = \lambda, n \geq 0$

2. MODELO LINEAL DE CRECIMIENTO CON INMIGRACION.

En este modelo interviene un factor externo, la inmigración. Suponga que cada individuo genera otro con tasa λ y que muere con tasa μ . Sea θ la tasa de inmigración. Nacimiento, muerte e inmigración son independientes para cada individuo y a los eventos pasados y futuros. Si se tienen $n \geq 0$ individuos en la población, entonces la tasa de nacimiento será $\lambda_n = n\lambda + \theta$, considerando como nacimiento todo lo que le de entrada a uno ó más miembros a la población. Si se tienen $n \geq 1$ individuos en la población, entonces la tasa de muerte será $\mu_n = n\mu$, pues cada individuo cuenta con esa tasa individual μ de muerte.

3. PROCESO DE NACIMIENTO PURO.

Un proceso de nacimiento puro es un proceso donde sólo ocurren "nacimientos", es decir que $\eta_n = 0$, para toda $n \geq 1$ y $\lambda_n > 0, n \geq 0$.

1.4.1 Ecuaciones diferenciales de Kolmogorov.

Recordando las probabilidades de transición en una cadena de Markov a tiempo continuo, se tiene

$$P_{ij}(t) = P(X_{t+s} = j | X_s = i) \quad (1.12)$$

Si se desea averiguar algo acerca de la posición de nuestros estados en el futuro, se puede aprovechar la propiedad de Markov. Se enuncian las ecuaciones de Chapman-Kolmogorov correspondientes a las cadenas de Markov continuas.

Proposición 1.4.5. Ecuaciones de Chapman-Kolmogorov

Para toda $s, t, i, j \in S$.

$$P_{ij}(t+s) = \sum_{k \in S} P_{ik}(t)P_{kj}(s). \quad (1.13)$$

Demostración. La demostración es similar a la presentada en el caso discreto, por lo que se omite. \square

Observe que en el caso discreto para el cálculo de la probabilidad de transición en n pasos se utilizan las ecuaciones de Chapman-Kolmogorov 1.2.6 de manera recursiva. Para el caso continuo, al usar las ecuaciones de Chapman-Kolmogorov se necesita de la probabilidad de transición expresada recientemente en la ecuación (1.12). Para calcular estas probabilidades se utilizan las ecuaciones diferenciales de Kolmogorov y para su presentación se recurre con fines prácticos al siguiente lema.

Lema 1.4.6. Para $i, j \in S$ se tiene:

1. $\lim_{t \rightarrow 0} \frac{1 - P_{ii}(t)}{t} = \nu_i$, tasa de retirada i .
2. $\lim_{t \rightarrow 0} \frac{P_{ij}(t)}{t} = q_{ij}$, $i \neq j$, tasa de transición de i a j .

Teorema 1.4.7. ECUACIONES BACKWARD KOLMOGOROV.

Sea $X = \{X_t : t \in [0, \infty)\}$ una cadena de Markov con espacio de estados S y matriz de transición $P(t) = (P_{ij}(t))_{i,j \in S}$. Entonces para toda $i, j \in S$ y $t \geq 0$.

$$P'_{ij}(t) = \sum_{k \neq i, k \in S} q_{ik}P_{kj}(t) - \nu_i P_{ij}(t)$$

Demostración. Se divide la demostración en dos partes. En la primera, se cuenta con S espacio de estados finito y en la segunda S con espacio de estados infinito numerable. Suponga S finito. Por las ecuaciones de Chapman-Kolmogorov de la proposición 1.4.5, se puede expresar,

$$P_{ij}(t+h) = \sum_{k \in S} P_{ik}(h)P_{kj}(t).$$

De esta igualdad se tiene que

$$\begin{aligned} P_{ij}(t+h) &= P_{ii}(h)P_{ij}(t) + \sum_{k \neq i} P_{ik}(h)P_{kj}(t) \\ &= P_{ii}(h)P_{ij}(t) + \sum_{k \neq i} P_{ik}(h)P_{kj}(t) + P_{ij}(t) - P_{ij}(t). \end{aligned}$$

Agrupando los términos convenientemente, resulta que

$$P_{ij}(t+h) = P_{ij}(t)[P_{ii}(h) - 1] + \sum_{k \neq i} P_{ik}(h)P_{kj}(t) + P_{ij}(t)$$

y por lo tanto,

$$P_{ij}(t+h) - P_{ij}(t) = \sum_{k \neq i} P_{ik}(h)P_{kj}(t) - (1 - P_{ii}(h))P_{ij}(t).$$

Dividiendo entre h , se obtiene,

$$\frac{P_{ij}(t+h) - P_{ij}(t)}{h} = \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) - \frac{(1 - P_{ii}(h))}{h} P_{ij}(t).$$

Finalmente, haciendo $h \rightarrow 0$ y usando el lema 1.4.6 y la definición de derivada se tiene que

$$P'_{ij}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t) - \nu_i P_{ij}(t)$$

Tome ahora S infinito numerable. En este caso se tiene que analizar la expresión

$$\lim_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t)$$

en dos casos por separado. Se tiene que demostrar que

$$\lim_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t).$$

Note que por definición de límite se tiene que

$$\liminf_{h \rightarrow 0} a_h \leq \lim_{h \rightarrow 0} a_h \leq \limsup_{h \rightarrow 0} a_h. \quad (1.14)$$

Si el $\lim_{h \rightarrow 0} a_h$ existe, entonces

$$\liminf_{h \rightarrow 0} a_h = \lim_{h \rightarrow 0} a_h = \limsup_{h \rightarrow 0} a_h.$$

De este modo, si se demuestra que

$$(i). \liminf_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) \geq \sum_{k \neq i} q_{ik} P_{kj}(t)$$

y que

$$(ii). \limsup_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) \leq \sum_{k \neq i} q_{ik} P_{kj}(t)$$

entonces usando la ecuación (1.14) con $a_h = \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t)$, se obtiene que

$$\sum_{k \neq i} q_{ik} P_{kj}(t) \leq \lim_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) \leq \sum_{k \neq i} q_{ik} P_{kj}(t)$$

y el resultado se sigue.

(i). Tome a $M \in S$ fijo y arbitrario, entonces

$$\begin{aligned} \sum_{k \neq i, k < M} q_{ik} P_{kj}(t) &= \liminf_{h \rightarrow 0} \sum_{k \neq i, k < M} \frac{P_{ik}(h)}{h} P_{kj}(t) \\ &\leq \liminf_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) \end{aligned}$$

dado que M es arbitrario, la expresión vale para todo M . De este modo,

$$\liminf_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) \geq \sum_{k \neq i, k < M} q_{ik} P_{kj}(t).$$

(ii). Tome ahora $M \in S$, arbitrario tal que $M > i$. Entonces

$$\begin{aligned} \limsup_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) &\stackrel{(1)}{=} \limsup_{h \rightarrow 0} \sum_{k \neq i, k < M} \frac{P_{ik}(h)}{h} P_{kj}(t) + \limsup_{h \rightarrow 0} \sum_{k \geq M} \frac{P_{ik}(h)}{h} P_{kj}(t) \\ &\stackrel{(2)}{\leq} \limsup_{h \rightarrow 0} \sum_{k \neq i, k < M} \frac{P_{ik}(h)}{h} P_{kj}(t) + \limsup_{h \rightarrow 0} \sum_{k \geq M} \frac{P_{ik}(h)}{h} \\ &\stackrel{(3)}{=} \limsup_{h \rightarrow 0} \sum_{k \neq i, k < M} \frac{P_{ik}(h)}{h} P_{kj}(t) + \limsup_{h \rightarrow 0} \left[\frac{1}{h} \left(1 - \sum_{k < M} P_{ik}(h) \right) \right] \\ &= \limsup_{h \rightarrow 0} \sum_{k \neq i, k < M} \frac{P_{ik}(h)}{h} P_{kj}(t) + \limsup_{h \rightarrow 0} \left[\frac{1}{h} \left(1 - P_{ii}(h) - \sum_{k \neq i, k < M} P_{ik}(h) \right) \right] \\ &\stackrel{(4)}{=} \sum_{k \neq i, k < M} \limsup_{h \rightarrow 0} \frac{P_{ik}(h)}{h} P_{kj}(t) + \limsup_{h \rightarrow 0} \frac{1 - P_{ii}(h)}{h} - \sum_{k \neq i, k < M} \limsup_{h \rightarrow 0} \frac{P_{ik}(h)}{h} \\ &\stackrel{(5)}{=} \sum_{k \neq i, k < M} q_{ik} P_{kj}(t) + \nu_i - \sum_{k \neq i, k < M} q_{ik} \end{aligned}$$

(1). Dividiendo la suma en $k \geq M$ y $k < M$. Note que $M > i$ y $k \geq M$, entonces $k \geq i$ y obviamente $k \neq i$.

(2). $P_{kj}(t) \leq 1$.

(3). $\sum_{k \geq M} P_{ik}(h) = 1 - \sum_{k < M} P_{ik}(h)$.

(4). Con suma finita cabe la posibilidad de intercambiar la suma y el límite.

(5). Lema 1.4.6.

De esta forma se tiene que para toda $M > i$ vale

$$\limsup_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) \leq \sum_{k \neq i, k < M} q_{ik} P_{kj}(t) + \nu_i - \sum_{k \neq i, k < M} q_{ik}$$

tomando el límite cuando M tiende a infinito, se obtiene

$$\begin{aligned} \limsup_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) &\leq \sum_{k \neq i, k \in S} q_{ik} P_{kj}(t) + \nu_i - \sum_{k \neq i, k \in S} q_{ik} \\ &\stackrel{(6)}{=} \sum_{k \neq i, k \in S} q_{ik} P_{kj}(t) + \nu_i - \nu_i \\ \limsup_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) &\leq \sum_{k \neq i} q_{ik} P_{kj}(t) \end{aligned}$$

$$(6). \nu_i = \sum_{j \neq i, j \in S} q_{ij}$$

De esta forma (i), (ii) y de la ecuación (1.14), sigue que

$$\lim_{h \rightarrow 0} \sum_{k \neq i} \frac{P_{ik}(h)}{h} P_{kj}(t) = \sum_{k \neq i} q_{ik} P_{kj}(t)$$

y por lo tanto

$$P'_{ij}(t) = \lim_{h \rightarrow 0} \frac{P_{ij}(t+h) - P_{ij}(t)}{h} = \sum_{k \neq i, k \in S} q_{ik} P_{kj}(t) - \nu_i P_{ij}(t)$$

con lo que queda la demostración terminada. □

Teorema 1.4.8. *ECUACIONES FORWARD KOLMOGOROV.*

Para toda i, j y $t \geq 0$

$$P'_{ij}(t) = \sum_{k \neq j} P_{ik}(t) q_{kj} - \nu_j P_{ij}(t).$$

Demostración. La demostración aunque similar adhiere otros supuestos más a los expuestos en la demostración anterior para establecer su validez.

Nota: Para una mayor información consultar Parzen, Feller, Karlin, Ross. \square

Una aplicación sencilla del teorema 1.4.7 (ó teorema 1.4.8), es el cálculo de $P_{ij}(t)$ cuando la cadena de Markov $X = \{X_t : t \in [0, \infty)\}$ tiene solamente dos estados, es decir $S = \{0, 1\}$. De este forma, se considera una cadena de Markov a tiempo continuo, con $S = \{0, 1\}$. El tiempo que pasa en cada estado antes de hacer una transición se distribuye exponencialmente de la siguiente manera. Si se encuentra en 0, se permanece en 0 por un espacio de tiempo exponencialmente distribuido con parámetro λ antes de cambiarnos a 1. Si se está en 1, se permanece en 1 por un espacio de tiempo exponencialmente distribuido con parámetro η antes de ir a 0. Por tanto, estableciendo las ecuaciones Forward Kolmogorov, se tiene para $i, j \in S$.

$$\begin{aligned} P'_{ij}(t) &= \sum_{k \neq j} P_{ik}(t)q_{kj} - \nu_j P_{ij}(t) \\ P'_{00}(t) &= P_{01}(t)q_{10} - \nu_0 P_{00}(t) \\ &= P_{01}(t)\eta - \lambda P_{00}(t) \\ P_{01}(t) + P_{00}(t) &= 1 \\ &= \eta(1 - P_{00}(t)) - \lambda P_{00}(t) \\ &= \eta - (\eta + \lambda)P_{00}(t) \end{aligned}$$

De donde resulta el siguiente conjunto de ecuaciones diferenciales

$$\begin{aligned} \text{(a).} \quad P'_{00}(t) &= \eta - (\eta + \lambda)P_{00}(t) \\ \text{(b).} \quad P'_{11}(t) &= \lambda - (\eta + \lambda)P_{11}(t) \end{aligned}$$

Las ecuaciones (a) y (b) pueden ser resueltas con una técnica muy sencilla. Se verá como resolver la ecuación (a). Para la ecuación (b) el proceso es similar con los cambios necesarios,

$$\begin{aligned}
 P'_{00}(t) + (\eta + \lambda)P_{00}(t) &= \eta \\
 \exp((\eta + \lambda)t)(P'_{00}(t) + (\eta + \lambda)P_{00}(t)) &= \eta \exp((\eta + \lambda)t) \\
 \frac{d}{dt}(P_{00}(t) \exp((\eta + \lambda)t)) &= \eta \exp((\eta + \lambda)t) \\
 \int_0^t \frac{d}{ds}(P_{00}(s) \exp((\eta + \lambda)s)) ds &= \int_0^t \eta \exp((\eta + \lambda)s) ds \\
 P_{00}(t) \exp((\eta + \lambda)t) - 1 &= \frac{\eta}{\eta + \lambda} \exp((\eta + \lambda)t) - \frac{\eta}{\eta + \lambda} \\
 P_{00}(t) &= \frac{\eta}{\eta + \lambda} + \frac{\lambda}{\eta + \lambda} \exp(-(\eta + \lambda)t)
 \end{aligned}$$

De manera análoga se tiene

$$P_{11}(t) = \frac{\lambda}{\eta + \lambda} + \frac{\eta}{\eta + \lambda} \exp(-(\eta + \lambda)t)$$

y basándose en que

$$\begin{aligned}
 P_{10}(t) &= 1 - P_{11}(t) \\
 P_{01}(t) &= 1 - P_{00}(t)
 \end{aligned}$$

se obtienen las expresiones para $P_{ij}(t)$, $i, j \in \{0, 1\}$.

1.4.2 Probabilidades límite para una cadena de Markov continua

Algunas notaciones necesarias son:

- (i). Sea H_i la distribución de tiempo que una cadena de Markov continua gasta en i antes de hacer una transición por θ_i su media.
- (ii). Sea T_{ii} el tiempo entre transiciones sucesivas al estado i .

Definición 1.4.9. Sea $Z \geq 0$, una variable aleatoria. Se dice que Z es "lattice", si existe $d \geq 0$, tal que $\sum_{n=0}^{\infty} P(Z = nd) = 1$. Es decir, Z es lattice si toma solamente valores múltiplos de d , para alguna $d \geq 0$. La d más grande que cumple lo anterior corresponde con $d(Z)$, período de Z .

Definición 1.4.10. Sea X'_n una variable aleatoria que registra el estado visitado, en el n -ésimo cambio de estado, por el proceso de Markov $X = \{X_t : t \geq 0\}$. Entonces $X' = \{X'_n, n = 0, 1, \dots\}$ es una cadena de Markov con espacio de estados S y matriz de transición $P = (P_{ij}), i, j \in S$ y X' es llamada **cadena de Markov anidada en X** .

Observación: Note que si X' es una cadena de Markov irreducible y X es una cadena de Markov anidada en X' , entonces X es irreducible.

Se enuncia el siguiente teorema sin presentar la demostración porque involucraría conceptos que no serían utilizados en lo restante de esta tesis.

Teorema 1.4.11. Sea $X = \{X_t : t \in [0, \infty)\}$ una cadena Markov continua irreducible, con espacio de estados S , T_{ii} con una distribución no-lattice, es decir, X es aperiódica y $X' = \{X'_n, n \geq 0\}$ la cadena de Markov anidada recurrente positiva y con distribución estacionaria $\pi(i), i \in S$, entonces

$$P(i) = \lim_{t \rightarrow \infty} P(X_t = i | X_0 = j) = \frac{\pi(i)\mu_i}{\sum_{j \in S} \pi(j)\mu_j} \quad (1.15)$$

independiente del estado inicial.

Nota: El resultado se encuentra en libros de procesos estocásticos. Se sugiere Ross, contenido en los capítulos: cadenas de Markov y cadenas de Markov a tiempo continuo.

Observación: Note que bajo las condiciones del teorema 1.4.11 se tiene que la distribución límite para $X = \{X_t : t \geq 0\}$ existe y es dada por la ecuación (1.15).

Proposición 1.4.12. Sea $X = \{X_t : t \in [0, \infty)\}$, una cadena de Markov continua satisfaciendo las condiciones del teorema 1.4.11. Entonces el límite

$$\lim_{t \rightarrow \infty} P(X_t = j | X_0 = i)$$

existe y es tal que

$$\lim_{t \rightarrow \infty} P_{ij}(t) = P(j), j \in S$$

con

$$P(j) = \frac{\frac{\pi(j)}{\nu_j}}{\sum_{i \in S} \frac{\pi(i)}{\nu_i}} \quad j \in S,$$

donde $\pi(\cdot)$ es la única solución no negativa de

$$\pi(j) = \sum_{i \in S} \pi(i)P_{ij}, \quad \sum_{i \in S} \pi(i) = 1.$$

Demostración. Es una aplicación directa del teorema 1.4.11, dado que

$$P(j) = \frac{\frac{\pi(j)}{\mu_j}}{\sum_i \frac{\pi(i)}{\mu_i}}$$

Una cadena de Markov a tiempo continuo es un proceso Semi-Markov con $F_{ij}(t) = 1 - \exp(-\nu_i t)$; de este modo, μ_i en el teorema 1.4.11 es tal que

$$\mu_i = \frac{1}{\nu_i}, \quad i \in S$$

y el resultado se sigue. □

Observación: Note que de la proposición 1.4.12 y de la definición de tasa de transición 1.4.4 se tiene que $P(i), i \in S$ es la solución del siguiente sistema de ecuaciones

$$\begin{aligned} \nu_j P(j) &= \sum_i P(i)q_{ij} \\ \sum_j P(j) &= 1. \end{aligned} \tag{1.16}$$

La interpretación de las ecuaciones (1.16) es la siguiente. La expresión $\nu_j P(j)$ corresponde a la tasa a la cual el proceso sale de j y $\sum_{i \neq j} P(i)q_{ij}$ representa a la tasa a la cual el proceso entra en j . Entonces las ecuaciones dicen que la tasa con que el proceso deja un estado es la misma con la cual entra en este mismo. De este modo ellas son el equivalente a las ecuaciones de balance total para el caso de cadenas de Markov discretas.

Definición 1.4.13. Cuando una cadena de Markov a tiempo continuo es irreducible y $P(j) > 0$, para toda $j \in S$, se dice que es *ergódica*.

Ejemplo: Proceso de nacimiento y muerte.

Sea $X = \{X_t : t \in [0, \infty)\}$ un proceso de nacimiento y muerte. Se tiene que la tasa de salida de un estado es equivalente a la tasa de entrada al estado. Por tanto, se considera el siguiente sistema,

$$\begin{aligned}\lambda_0 P(0) &= \eta_1 P(1) \\ \eta_n P(n) + \lambda_n P(n) &= \eta_{n+1} P(n+1) + \lambda_{n-1} P(n-1), \text{ con } n \geq 1.\end{aligned}$$

La solución de este sistema es dada por,

$$P(n) = \frac{\lambda_{n-1} \lambda_{n-2} \cdots \lambda_0}{\eta_n \eta_{n-1} \cdots \eta_1} P(0), \quad n \geq 0$$

Esto se demuestra por inducción

$$\begin{aligned}\lambda_0 P(0) &= \eta_1 P(1) \\ \eta_n P(n) + \lambda_n P(n) &= \eta_{n+1} P(n+1) + \lambda_{n-1} P(n-1)\end{aligned} \tag{1.17}$$

para $n = 1$, se tiene de las ecuaciones (1.17)

$$P(1) = \frac{\lambda_0}{\eta_1} P(0), \quad P(2) = \frac{\lambda_0 \lambda_1}{\eta_1 \eta_2} P(0)$$

suponga que sea valido para $k \leq n$, es decir

$$P(k) = \frac{\lambda_0 \lambda_1 \cdots \lambda_{k-1}}{\eta_1 \eta_2 \cdots \eta_k} P(0) \tag{1.18}$$

y se demostrará la validez para $k + 1$. De las ecuaciones anteriores (1.17) con $n = k$

$$\begin{aligned}
\eta_{k+1}P(k+1) + \lambda_{k-1}P(k-1) &= (\lambda_k + \eta_k)P(k) \\
P(k+1) &= \frac{(\lambda_k + \eta_k)}{\eta_{k+1}}P(k) - \frac{\lambda_{k-1}}{\eta_{k+1}}P(k-1) \\
&\stackrel{(1)}{=} \frac{\lambda_k + \eta_k}{\eta_{k+1}} \frac{\lambda_0 \cdots \lambda_{k-1}}{\eta_1 \cdots \eta_k} P(0) - \frac{\lambda_{k-1}}{\eta_{k+1}} \frac{\lambda_0 \cdots \lambda_{k-2}}{\eta_1 \cdots \eta_{k-1}} P(0) \\
&= \frac{\lambda_k + \eta_k}{\eta_{k+1}} \frac{\lambda_0 \cdots \lambda_{k-1}}{\eta_1 \cdots \eta_k} P(0) - \frac{\lambda_0 \cdots \lambda_{k-2} \lambda_{k-1}}{\eta_1 \cdots \eta_{k-1} \eta_{k+1}} P(0) \\
&= \frac{\lambda_0 \cdots \lambda_{k-1}}{\eta_1 \cdots \eta_k} \frac{P(0)}{\eta_{k+1}} \left[\lambda_k + \eta_k - \eta_k \right] \\
&= \frac{\lambda_0 \cdots \lambda_{k-1}}{\eta_1 \cdots \eta_k} \frac{P(0)}{\eta_{k+1}} \lambda_k \\
P(k+1) &= \frac{\lambda_0 \cdots \lambda_{k-1} \lambda_k}{\eta_1 \cdots \eta_k \eta_{k+1}} P(0)
\end{aligned}$$

(1). Por la ecuación (1.18). Sea que por la hipótesis de inducción $p(\cdot)$ tiene que satisfacer

$$\sum_{n=0}^{\infty} P(n) = 1,$$

es decir,

$$\sum_{n=1}^{\infty} P(n) + P(0) = 1$$

entonces

$$\begin{aligned}
P(0) + \sum_{n=1}^{\infty} \frac{\lambda_0 \cdots \lambda_{n-2} \lambda_{n-1}}{\eta_1 \cdots \eta_{n-1} \eta_n} P(0) &= 1 \\
P(0) \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_{n-1} \lambda_{n-2} \cdots \lambda_0}{\eta_n \eta_{n-1} \cdots \eta_1} \right) &= 1 \\
P(0) &= \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_{n-1} \lambda_{n-2} \cdots \lambda_0}{\eta_n \eta_{n-1} \cdots \eta_1} \right)^{-1}
\end{aligned}$$

y

$$P(n) = \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\eta_n\eta_{n-1}\cdots\eta_1} \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\eta_n\eta_{n-1}\cdots\eta_1}\right)^{-1}$$

para $n \geq 1$.

Para que las probabilidades límites sean estrictamente positivas se debe de cumplir como condición que

$$\sum_{n=1}^{\infty} \frac{\lambda_{n-1}\lambda_{n-2}\cdots\lambda_0}{\eta_n\eta_{n-1}\cdots\eta_1} < \infty.$$

Capítulo 2

Acoplamiento y cadenas de Markov

“The proper business of probabilists is calculating probabilities. Often exact calculations are tedious or impossible, so we resort to approximation.”

D. Aldous.

Probability approximations via the Poisson Clumping Heuristic.

2.1 Introducción

Este capítulo tiene como objetivo exponer el uso de la **técnica de acoplamiento** para estimar el tiempo de llegada al estado estacionario de una cadena de Markov ergódica. Este tiempo para llegar al estado estacionario, **tiempo de corte**.

El **tiempo de corte** para una cadena de Markov ergódica es el tiempo n^* , tal que, para $n \geq n^*$ se tiene a $P(X_n = \cdot)$ distribución muy cerca de la distribución estacionaria $\pi(\cdot)$, de modo que se puede considerar que la cadena se encuentra en el estado estacionario y usar las distribuciones para obtener información sobre el fenómeno modelado por esta(s) cadena(s) de Markov.

La **distancia de variación total** y la **distancia de separación total** serán definidos en la siguiente sección siendo los medios utilizados para medir la diferencia entre $P(X_n)$ y π .

El acoplamiento tiene un gran campo de acción. En las siguientes páginas, el acoplamiento para cadenas de Markov obtendrá una cota superior para la

distancia de separación total y la distancia de variación total con ello se estimará el tiempo n^* que hace a esta cota suficientemente pequeña. Entre diversos métodos semejantes en objetivo al acoplamiento se pueden mencionar *tiempos fuertes uniformes, análisis de Fourier y ordenamiento estocástico*.

Serán presentados los resultados para una cadena de Markov del tipo discreto, aclarando que resultados semejantes valen para una cadena del tipo continuo bajo características particulares.

2.2 Distancia entre distribuciones

El trabajo a realizar establece de manera natural algún concepto que otorgue un parámetro que establezca “cercanía” o “lejanía” entre dos distribuciones cualesquiera. En nuestra posición es inmediato pensar en medir la distancia entre distribuciones a usar: la distribución de la cadena de Markov al tiempo n (ó t) y la distribución de equilibrio (o estacionaria). Esto explica la existencia de este apartado.

2.2.1 Distancia de variación total

La distancia de variación total ayudará a medir la distancia que separa dos distribuciones de probabilidad definidas sobre un mismo espacio de estados de un espacio común de probabilidad.

Definición 2.2.1. Sea S un espacio de estados, $A \subseteq S$ y $Q_i, i = 1, 2$ distribuciones cualesquiera sobre S . Entonces se llama **distancia de variación total** a

$$\begin{aligned} d(Q_1, Q_2) &= \|Q_1 - Q_2\| \\ &= \frac{1}{2} \left(\sum_{i \in S} |Q_1(i) - Q_2(i)| \right) \\ &\stackrel{(1)}{=} \max_{ACS} |Q_1(A) - Q_2(A)|. \end{aligned}$$

(1) Ver referencia en Aldous y Diaconis (1986)

Dado que la definición 2.2.1 es aplicable a dos distribuciones cualesquiera, se puede reformular esta definición a un caso con ciertas características específicas.

Definición 2.2.2. Sea P , una matriz de transición ergódica de una cadena de Markov $X = \{X_n : n = 0, 1, 2, \dots\}$ con S espacio de estados. Sea i_0 un estado inicial cualquiera y π la distribución estacionaria única. Sea indicado por $P_{i_0}(X_n)$ la distribución al tiempo n de X dado que $X_0 = i_0$, es decir, $P_{i_0}(X_n = j) = P(X_n = j | X_0 = i_0)$, entonces la **distancia de variación total** entre la distribución al tiempo n (ó t) y la distribución estacionaria es:

$$d_{i_0}(n) = |P_{i_0}(X_n) - \pi|$$

Sea definida además,

$$d(n) = \max_{i_0 \in S} \{d_{i_0}(n)\}$$

Observaciones:

1. Suponga que inicialmente los estados en el espacio en el cual se trabaja, tienen igual probabilidad, es decir, el espacio es equiprobable, se puede asegurar que,

$$d_{i_0}(t) = d(t), \text{ para todo } i_0 \in S,$$

pues no hay dependencia explícita sobre el estado del cual parte la cadena X . Haciendo referencia a las distribuciones estacionarias, se tenía que $P_{ij}(n) \rightarrow \pi(j)$, cuando $n \rightarrow \infty$. De este modo

$$d_{i_0}(n) = |P_{i_0}(X_n) - \pi| \rightarrow 0, \quad \text{cuando } n \rightarrow \infty$$

es decir $d(n) \rightarrow 0$, cuando $n \rightarrow \infty$.

2. Note que de la definición de $d(n)$ se da naturalmente que $0 \leq d(n) \leq 1$, asumiendo que $d(n) \geq 0$. Ahora se tiene que por definición:

$$\begin{aligned} d_{i_0}(n) &= \frac{1}{2} \sum_{i \in S} |P_{i_0 i}^{(n)} - \pi(i)| \\ &\stackrel{(1)}{\leq} \frac{1}{2} \left[\sum_{i \in S} |P_{i_0 i}^{(n)}| + \sum_{i \in S} |\pi(i)| \right] \\ &\stackrel{(2)}{=} 1. \end{aligned}$$

(1). Por la desigualdad del triángulo $|a + b| \leq |a| + |b|$ y del hecho que las funciones de probabilidad son positivas.

(2). Del hecho que $\sum_{i \in S} P_{ij}^{(n)} = 1$ y $\sum_{i \in S} \pi(i) = 1$.

3. Se tiene también que $d(n)$ es una métrica. Esta afirmación es fácilmente demostrada dado que para cualesquiera Q_1 y Q_2 en un mismo espacio de estados.

i. $d(Q_1, Q_2) = d(Q_2, Q_1)$, dado que,

$$\|Q_1(A) - Q_2(A)\| = \|(-1)[Q_2(A) - Q_1(A)]\| = \|Q_2(A) - Q_1(A)\|$$

ii. $d(Q_1, Q_2) = 0$, si y sólo si $Q_1(A) = Q_2(A)$ para todo $A \subset S$, es decir, si y sólo si $Q_1 \equiv Q_2$

iii.

$$\begin{aligned} d(Q_1, Q_3) &= \frac{1}{2} \sum_{i \in S} |Q_1(i) - Q_3(i)| \\ &= \frac{1}{2} \sum_{i \in S} |Q_1(i) - Q_2(i) + Q_2(i) - Q_3(i)| \\ &\stackrel{(1)}{\leq} \frac{1}{2} \sum_{i \in S} [|Q_1(i) - Q_2(i)| + |Q_2(i) - Q_3(i)|] \\ &= \frac{1}{2} \sum_{i \in S} |Q_1(i) - Q_2(i)| + \frac{1}{2} \sum_{i \in S} |Q_2(i) - Q_3(i)| \\ &= d(Q_1, Q_2) + d(Q_2, Q_3) \end{aligned}$$

para toda $Q_i, i = 1, 2, 3$ definidas en el mismo espacio de estados.

(1). Desigualdad del triángulo $|a + b| \leq |a| + |b|$

Se establece en el capítulo 1, en la sección de cadenas a tiempo reversible, en la definición 1.3.2, que cuando $P_{ij} = P_{ij}^*$, para toda $i, j \in S$, la cadena de Markov X se dice reversible en el tiempo con respecto a su distribución estacionaria π .

La relación de este hecho con la distancia de variación total queda expuesta en el siguiente lema.

Lema 2.2.3. *Sea $d(n), (d^*(n))$ la distancia de variación total para una cadena de Markov X reversible en el tiempo, (para su cadena de Markov reversa X^*). Entonces*

$$d(n) = d^*(n)$$

Demostración. Sea S el espacio de estados de $X = \{X_n : n = 0, 1, \dots\}$ y sean $P^{(n)} = (P_{ij}^{(n)})_{i,j \in S}$ y $P^{*(n)} = (P_{ij}^{*(n)})_{i,j \in S}$, las matrices de transición de n pasos para X y su cadena reversa X^*

$$d(n) = \sum_{j \in S} |P_{ij}^{(n)} - \pi| = \sum_{j \in S} |P_{ij}^{*(n)} - \pi| = d^*(n)$$

De donde se establece que si la cadena de Markov es reversible, basta analizar una cota superior, ya sea para $d(n)$ ó para $d^*(n)$. \square

2.2.2 Distancia de separación total

Definición 2.2.4. Sean Q_1, Q_2 dos distribuciones de probabilidad cualesquiera en un espacio de estados S . Será llamada distancia de separación total a:

$$s(Q_1, Q_2) = \max_{i \in S} \left(1 - \frac{Q_1(i)}{Q_2(i)} \right).$$

Observaciones:

- i. Note que $0 \leq s(Q_1, Q_2) \leq 1$.
- ii. También se tiene que $s(Q_1, Q_2)$ no es una métrica. Esto se verifica al notar que

$$s(Q_1, Q_2) = \max_{i \in S} \left(1 - \frac{Q_1(i)}{Q_2(i)} \right)$$

no siempre es igual

$$s(Q_2, Q_1) = \max_{i \in S} \left(1 - \frac{Q_2(i)}{Q_1(i)} \right)$$

para Q_1 y Q_2 probabilidades en un mismo espacio de estados S .

Análogamente a la definición de distancia de variación total (2.2.1), se puede particularizar dando la siguiente definición.

Definición 2.2.5. Sea P una matriz de transición ergódica de una cadena de Markov, con S espacio de estados. Sea i_0 un estado inicial cualquiera y π la distribución estacionaria única. Sea $P_{i_0}^{(n)}(X_n)$, la distribución al tiempo

n de X dado que $X_0 = i_0$. Sean $Q_1 = P_{i_0}(X_n)$ y $Q_2 = \pi$, se tiene que la **distancia de separación total** es:

$$s_{i_0}(n) = \max_{i \in S} \left(1 - \frac{P_{i_0}(X_n = i)}{\pi(i)} \right) = \max_{i \in S} \left(1 - \frac{P_{i_0 i}^{(n)}}{\pi(i)} \right)$$

También se define:

distancia máxima de separación por,

$$s^*(n) = \max_{i_0 \in S} s_{i_0}(n) = \max_{i_0 \in S} \left(1 - \frac{P_{i_0 i}^{(n)}}{\pi(i)} \right)$$

Observación:

1. Cuando se tiene un espacio en el cual todos los estados cuentan con la misma probabilidad no se depende del estado inicial, por tanto $s^*(n) = s_{i_0}(n)$, para todo $i_0 \in S$.

2. Cuando se tiene una cadena de Markov a tiempo continuo las definiciones de distancia de variación y separación total son las mismas, solamente la notación cambia de $s_{i_0}(n)$ a $s_{i_0}(t)$, $s(n)$ a $s(t)$ y de $d_{i_0}(n)$ a $d_{i_0}(t)$, $d(n)$ a $d(t)$. Dado que en las aplicaciones se va a tener una cadena discreta, se usará la notación para tiempo discreto.

Proposición 2.2.6. Para algunas Q_1, Q_2 distribuciones en un mismo espacio de estados S , sucede que,

$$d(Q_1, Q_2) \leq s(Q_1, Q_2)$$

Demostración. Sea S el espacio de estados. Defina

$$B = \{i \in S : Q_1(i) < Q_2(i)\}$$

entonces el complemento de B es

$$B^c = \{i \in S : Q_1(i) \geq Q_2(i)\}.$$

De este modo, se escribe

$$\begin{aligned}
 d(Q_1, Q_2) &\stackrel{(1)}{=} \frac{1}{2} \sum_{i \in S} |Q_1(i) - Q_2(i)| \\
 &= \frac{1}{2} \sum_{i \in B} |Q_1(i) - Q_2(i)| + \frac{1}{2} \sum_{i \in B^c} |Q_1(i) - Q_2(i)| \\
 &= \frac{1}{2} \left[\sum_{i \in B} (Q_2(i) - Q_1(i)) + \sum_{i \in B^c} (Q_1(i) - Q_2(i)) \right] \\
 &\stackrel{(2)}{=} \frac{1}{2} \left[\sum_{i \in B} (Q_2(i) - Q_1(i)) + \sum_{i \in B} (1 - Q_1(i) - (1 - Q_2(i))) \right] \\
 &= \frac{1}{2} \left[\sum_{i \in B} (Q_2(i) - Q_1(i)) + \sum_{i \in B} (Q_2(i) - Q_1(i)) \right] \\
 &= \sum_{i \in B} (Q_2(i) - Q_1(i)) \\
 &= \sum_{i \in B} Q_2(i) \left(1 - \frac{Q_1(i)}{Q_2(i)}\right) \\
 &\leq \sum_{i \in B} Q_2(i) \max_{k \in S} \left(1 - \frac{Q_1(k)}{Q_2(k)}\right) \\
 &= \max_{k \in S} \left(1 - \frac{Q_1(k)}{Q_2(k)}\right) \sum_{i \in B} Q_2(i) \\
 &\stackrel{(3)}{\leq} \max_{k \in S} \left(1 - \frac{Q_1(k)}{Q_2(k)}\right) \\
 &= s(Q_1, Q_2)
 \end{aligned}$$

(1). Definición de $d(Q_1, Q_2)$.

(2). Dado que $P(A^c) = 1 - P(A)$ para A un subconjunto del espacio de estados donde P esta definida.

(3). Dado que $\sum_{i \in B} Q_1(i) \leq 1$. □

2.3 Tiempo aleatorio y tiempo de paro

Al hablar de acoplamiento se verá que un concepto fundamental será el tiempo de acoplamiento, es decir, el tiempo que tarda para acoplar los procesos que se estarán estudiando. Es de interés que este tiempo de acoplamiento sea finito. También se estudiará que este tiempo de acoplamiento es una variable aleatoria que cumple con ciertas propiedades. Son exactamente estas propiedades que se estudiarán en esta sección.

Definición 2.3.1. *La variable aleatoria T con valores en $\{0, 1, 2, \dots, \infty\}$, es un tiempo aleatorio con respecto a $\{Y_n\}$, si para cada $n = 0, 1, \dots$, el evento $\{T = n\}$ es determinado por $\{Y_0, Y_1, \dots, Y_n\}$, es decir, conociendo el conjunto de observaciones Y_0, Y_1, \dots, Y_n , se indica si existe la igualdad $T = n$ o no.*

Esto significa que la función indicadora del evento $\{T = n\}$ es una función de Y_0, Y_1, \dots, Y_n .

Será nombrado un tiempo de paro a aquel en el que se decide que una actividad se detenga. Se evaluará ese tiempo en conjunción a los intereses buscados.

Definición 2.3.2. *Si T es tiempo aleatorio con respecto a alguna sucesión de variables aleatorias $\{Y_n\}$ y $P(T < \infty) = 1$, entonces T es un tiempo de paro.*

Observación:

1. Una forma de notación de los tiempos de paro utilizada es la siguiente:

$$I_{(T=n)} = I_{(T=n)}(Y_0, \dots, Y_n) = \begin{cases} 1, & \text{si } T = n \\ 0, & \text{si } T \neq n. \end{cases}$$

Con esto se determina $\{T \geq n\}$, $\{T > n\}$, $\{T < n\}$, $\{T \leq n\}$. Por ejemplo,

$$I_{(T \leq n)} = \sum_{k=0}^n I_{(T=k)}(Y_0, \dots, Y_n)$$

$$I_{(T > n)} = 1 - I_{(T \leq n)}$$

2. Algunos ejemplos de tiempos de paro son los siguientes:

(i). El tiempo constante, $T = k$, para toda (Y_0, \dots, Y_n)

$$I_{(T=n)}(Y_0, \dots, Y_n) = \begin{cases} 0, & \text{si } n \neq k \\ 1, & \text{si } n = k \end{cases}$$

(ii). La primera vez que el proceso ingresa a un subconjunto $T(A) = \min\{n | Y_n \in A\}$

$$I_{(T(A)=n)}(Y_0, \dots, Y_n) = \begin{cases} 0, & \text{en otro caso} \\ 1, & \text{si } Y_n \in A \end{cases}$$

Propiedades: Algunas de las propiedades de los tiempos de paro son,

- i. Si S y T son tiempos de paro, entonces $S + T$ es un tiempo de paro dado que $I_{(S+T=n)} = \sum_{k=0}^n I_{(S=k)} I_{(T=n-k)}$.
- ii. El más pequeño de dos tiempos de paro S, T denotado como $S \wedge T = \min\{S, T\}$, es un tiempo de paro, dado que $I_{(S \wedge T > n)} = I_{(S > n)} I_{(T > n)}$.
- iii. El más grande de dos tiempos de paro, conservando la notación $S \vee T = \max\{S, T\}$, es un tiempo de paro, dado que $I_{(S \vee T \leq n)} = I_{(S \leq n)} I_{(T \leq n)}$.

Las consideraciones hechas desde el principio de esta sección permiten definir:

Definición 2.3.3. Sean $X = (X_n, n \geq 0)$ y $X' = (X'_n, n \geq 0)$ dos procesos cualesquiera en un mismo espacio de probabilidad. Un tiempo de paro, T es llamado **tiempo de acoplamiento** si,

$$T = \inf\{n : X'_n = X_n\}.$$

Este concepto abre paso a la siguiente sección.

2.4 Método de acoplamiento

Lo que se estudia en esta sección es el caso de acoplamiento entre cadenas de Markov. La importancia del acoplamiento reside en que es uno de los métodos de mayor uso dentro del campo de la probabilidad. (Con el fin de estudiar otros ejemplos de aplicación, referirse a Lindvall (1992)).

El método de acoplamiento consiste en la construcción de un espacio de probabilidad bidimensional con base en dos espacios de probabilidad unidimensionales.

La técnica de acoplamiento nace con Doeblin (1938). En la década de los 70's hubo un gran avance en los resultados de acoplamiento en el tema de cadenas de Markov. Desde esta época varios trabajos (teóricos y aplicados) fueron desarrollados con esta técnica, por ejemplo, Griffeath (1975), Thorrisson (1986), Diaconis (1992), entre otros. Se mostrará inicialmente la definición de acoplamiento entre dos versiones de una cadena de Markov con espacio de estados S y matriz de transición P . De este modo se tiene:

Definición 2.4.1. Sea $X = \{X_n, n \geq 0\}$ un proceso de Markov con espacio de estados S y matriz de transición P . Se fija $i_1, i_2 \in S$ estados, X^1 es la versión del proceso X dado que $X_0 = i_1$ y X^2 es la versión del proceso X dado que $X_0 = i_2$. Se construye un proceso

$$(X^1, X^2) = \{(X_n^1, X_n^2) : n = 0, 1, 2, \dots\}$$

de dimensión 2 y sea $T = \inf\{n | X_n^1 = X_n^2\}$.

Note que, $X_n^1 = X_n^2$ en $\{n \geq T\}$. El proceso (X^1, X^2) es llamado **acoplamiento y T es el tiempo de acoplamiento**.

Observaciones:

1. Note que para cada $n = 0, 1, \dots$ se tiene que X_n^1 se distribuye como X , con la cadena empezando en i_1 y X_n^2 se distribuye como X , con esta cadena empezando en i_2 .

Se muestra a continuación un resultado que será fundamental dentro de la aplicación del acoplamiento para acotar superiormente la distancia de variación total. El mismo resultado es válido para $s(n)$.

Teorema 2.4.2. Sean $X^1 = \{X_n^1 : n = 0, 1, \dots\}$ y $X^2 = \{X_n^2 : n = 0, 1, \dots\}$, procesos estocásticos y sus distribuciones de probabilidad respectivas $Q_i, i = 1, 2$, entonces para (X^1, X^2) un acoplamiento entre los dos procesos

$$d(n) \leq P(X_n^1 \neq X_n^2).$$

Dadas Q_1, Q_2 es posible construir (X^1, X^2) , tal que,

$$d(n) = P(X_n^1 \neq X_n^2)$$

Demostración. Se demostrará solamente la desigualdad de acoplamiento. Para demostrar la igualdad, se tendría que utilizar técnicas que van más allá de alcance de esta tesis, se sugiere para la consulta de esta demostración Thorisson y Griffeath.

Por definición se tiene que

$$d(Q_1, Q_2) = \max_{ACS} |Q_1(X_n^1 \in A) - Q_2(X_n^2 \in A)|$$

Tome $A \subset S$ arbitrario, entonces se divide en dos casos:

(i). $Q_1(X_n^1 \in A) \geq Q_2(X_n^2 \in A)$

$$\begin{aligned} |Q_1(X_n^1 \in A) - Q_2(X_n^2 \in A)| &= Q_1(X_n^1 = j) - Q_2(X_n^2 = j) \\ &= P(X_n^1 = j, X_n^1 = X_n^2) + P(X_n^1 = j, X_n^1 \neq X_n^2) \\ &\quad - P(X_n^2 = j, X_n^1 = X_n^2) - P(X_n^2 = j, X_n^1 \neq X_n^2) \\ &\stackrel{(1)}{=} P(X_n^1 = j, X_n^1 \neq X_n^2) - P(X_n^2 = j, X_n^1 \neq X_n^2) \\ &\leq P(X_n^1 = j, X_n^1 \neq X_n^2) \\ &= P(X_n^1 = j | X_n^1 \neq X_n^2) P(X_n^1 \neq X_n^2) \\ &\leq P(X_n^1 \neq X_n^2) \end{aligned}$$

(1). Pues (X^1, X^2) es un acoplamiento.

(ii). $Q_1(X_n^1 \in A) < Q_2(X_n^2 \in A)$. Análogo al procedimiento anterior. \square

Corolario 2.4.3. Sean $X^1 = \{X_n^1, n \geq 0\}$ y $X^2 = \{X_n^2, n \geq 0\}$, procesos estocásticos, con espacio de estados S y $Q_i, i = 1, 2$ sus distribuciones de probabilidad respectivas. Si T es el tiempo de acoplamiento y (X^1, X^2) es un acoplamiento, entonces

$$d(n) \leq P(T > n).$$

Demostración. Del teorema 2.4.2 se obtiene que

$$d(n) \leq P(X_n^1 \neq X_n^2)$$

Dado que (X^1, X^2) es un acoplamiento y T es el tiempo de acoplamiento, por la definición de tiempo de acoplamiento se tiene que $X_n^1 \neq X_n^2$ para $n < T$ (en otro caso se tendría $X_n^1 = X_n^2$). De este modo la relación entre eventos

$$\{X_n^1 \neq X_n^2\} \subseteq \{n < T\}$$

vale. De allí,

$$P(X_n^1 \neq X_n^2) \leq P(n < T)$$

y el resultado sigue. □

Al hacer la notación de (X^1, X^2) se estará implícitamente considerando un acoplamiento entre los procesos X^1 y X^2 . Para $X = (X_n, n \geq 0)$, un proceso estocástico con espacio de estados S , $P_i(X_n \in \cdot)$ la distribución de probabilidad del proceso X , al tiempo n dado que X que inicia en el estado i . Usando la notación $\rho_{ij}(n)$ y $\rho(n)$ para indicar,

$$\begin{aligned} \rho_{ij}(n) &= \|P_i(X_n \in \cdot) - P_j(X_n \in \cdot)\| \\ \rho(n) &= \max_{i,j \in S} \rho_{ij}(n). \end{aligned}$$

Con esta notación se tiene que en el siguiente lema en el punto (iii) se demuestra que la distancia de variación total es decreciente en el tiempo.

Lema 2.4.4. *Sea $X = \{X_n : n = 0, 1, \dots\}$ una cadena de Markov. Usando la definición 2.2.2 y las de $\rho_{ij}(n)$, $\rho(n)$ arriba presentadas, se tiene que:*

(i). $\rho(n) \leq 2d(n)$.

(ii). ρ es submultiplicativo, $\rho(n+m) \leq \rho(n)\rho(m)$.

(iii). $d(n)$ es decreciente.

Demostración. (i). Sean

$$\rho_{ij}(n) = \|P_i(X_n \in \cdot) - P_j(X_n \in \cdot)\|.$$

Sumando y sustrayendo $\pi(\cdot)$ y usando la desigualdad del triángulo, se tiene

$$\begin{aligned} \rho_{ij}(n) &= \|[P_i(X_n \in \cdot) - \pi(\cdot)] + [\pi(\cdot) - P_j(X_n \in \cdot)]\| \\ &\leq \|P_i(X_n \in \cdot) - \pi(\cdot)\| + \|P_j(X_n \in \cdot) - \pi(\cdot)\|. \end{aligned}$$

Tomando el máximo sobre $i, j \in S$

$$\max_{i,j} \rho_{i,j}(n) \leq \max_{i,j \in S} \left\{ \|(P_i(X_n \in \cdot) - \pi(\cdot))\| + \|(P_j(X_n \in \cdot) - \pi(\cdot))\| \right\}. \quad (2.1)$$

Dado que para dos funciones f, g definidas en un mismo dominio D y tales que $f(x), g(x) \geq 0$, para toda $x \in D$, se tiene que por la ecuación (2.1)

$$\max_{x \in D} [f(x) + g(x)] = \max_{x \in D} f(x) + \max_{x \in D} g(x),$$

entonces por definición de $d_i(n)$ y $d(n)$ y definición de $\rho(t)$ se tiene que

$$\rho(n) \leq 2d(n)$$

como se quería demostrar.

(ii). Usando el teorema 2.4.2, se fija i_1, i_2, m, n . Se construye un acoplamiento (X_m^1, X_m^2) , tal que

$$P(X_m^1 \in \cdot) = P(X_m \in \cdot | X_0 = i_1)$$

y

$$P(X_m^2 \in \cdot) = P(X_m \in \cdot | X_0 = i_2)$$

y tal que

$$\rho_{i_1, i_2}(m) = P(X_m^1 \neq X_m^2).$$

Sean

$$A_j = \{X_m^1 = j, X_m^2 = j\}$$

y

$$A_{jk} = \{X_m^1 = j, X_m^2 = k\}, j \neq k$$

construya (X_{m+n}^1, X_{m+n}^2) , tal que en los conjuntos A_j se tiene que

$$(a). X_{m+n}^1 = X_{m+n}^2$$

$$(b). P(X_{m+n}^1 \in \cdot | A_j) = P_j(X_n \in \cdot).$$

y en los conjuntos $A_{j,k}$, $k \neq j$, se tiene ,

$$\begin{aligned} (c). & P(X_{m+n}^1 \in \cdot | A_{j,k}) = P_j(X_n \in \cdot) \\ (d). & P(X_{m+n}^2 \in \cdot | A_{j,k}) = P_k(X_n \in \cdot) \\ (e). & P(X_{m+n}^1 \neq X_{m+n}^2 | A_{j,k}) = \rho_{jk}(n) \end{aligned}$$

Por la definición de $\rho(n)$ se tiene que el punto (e) equivale

$$P(X_{m+n}^1 \neq X_{m+n}^2 | A_{j,k}) = \rho_{j,k}(n) \leq \rho(n).$$

Por construcción del acoplamiento (X^1, X^2) se tiene que

$$\begin{aligned} P(X_{m+n}^1 \in \cdot) &= P(X_{m+n} \in \cdot | X_0 = i_1) \\ P(X_{m+n}^2 \in \cdot) &= P(X_{m+n} \in \cdot | X_0 = i_2) \end{aligned}$$

de donde

$$\begin{aligned} \rho_{i_1 i_2}(m+n) &= \|P_{i_1}(X_{m+n} \in \cdot) - P_{i_2}(X_{m+n} \in \cdot)\| \\ &\stackrel{(1)}{\leq} P(X_{m+n}^1 \neq X_{m+n}^2) \end{aligned}$$

para j fija

$$\begin{aligned} &\stackrel{(2)}{=} \sum_{j \in S} \sum_{k \neq j} P(X_{m+n}^1 \neq X_{m+n}^2 | A_{j,k}) P(X_m^1 = j, X_m^2 = k) \\ &= \sum_{j \in S} \sum_{k \neq j} \rho_{j,k}(n) P(X_m^1 = j, X_m^2 = k) \\ &\stackrel{(3)}{\leq} \rho(n) \sum_{j \in S} \sum_{k \neq j} P(X_m^1 = j, X_m^2 = k) \\ &= \rho(n) P(X_m^1 \neq X_m^2) \\ &\stackrel{(4)}{=} \rho(n) \rho_{i_1 i_2}(m) \\ &\leq \rho(n) \max_{i_1, i_2 \in S} \rho_{i_1 i_2}(m) \\ &= \rho(m) \rho(n). \end{aligned}$$

- (1). Desigualdad de acoplamiento.
- (2). Usando probabilidad condicional en $A_{j,k}$.
- (3). $\rho_{jk}(t) \leq \max_{j,k} \rho_{j,k}(t) = \rho(t)$.

(4). De la construcción del acoplamiento.

(iii). Use el acoplamiento construido en el punto (ii), pero ahora con $P(X_m^2 \in \cdot) = \pi(\cdot)$ distribución estacionaria de la cadena de Markov y tomando

$$P(X_m^1 \neq X_m^2) = d_{i_1}(m)$$

en lugar de

$$P(X_m^1 \neq X_m^2) = \rho_{i_1 i_2}(m).$$

Se tiene que $\rho_{i_1 i_2}(m+n) = d_{i_1}(m+n)$.

Por lo tanto, sigue que

$$d_{i_1}(m+n) \leq P(X_{m+n}^1 \neq X_{m+n}^2) \stackrel{(1)}{\leq} \rho(n)P(X_m^1 \neq X_m^2) = \rho(n)d_{i_1}(m) \stackrel{(2)}{\leq} d_{i_1}(m).$$

(1). Por el inciso (ii).

(2). Por $\rho(n) \leq 1$. □

A continuación se expone como utilizar el **método de acoplamiento para acotar superiormente la distancia de variación total** en el caso de cadenas de Markov ergódicas.

Sea $X = (X_n, n \geq 0)$, una cadena de Markov con espacio de estados $S = \{0, 1, 2, \dots\}$, P su matriz de transición aperiódica, irreducible y recurrente positiva. Bajo estas condiciones se tiene por la observación de la página 22 que X se aproxima al estado estacionario cuando $n \rightarrow \infty$, es decir,

$$P(X_n = j) = \sum_{i \in S} \pi_0(i) P_{ij}^{(n)} \rightarrow \pi(j) \text{ cuando } n \rightarrow \infty$$

donde $\pi(\cdot)$ es distribución estacionaria única de X , $P^{(n)}$ es la matriz en el n -ésimo paso y π_0 la distribución inicial de X .

Definición 2.4.5. Acoplamiento para cadenas de Markov.

Sea P matriz de transición sobre el espacio de estados S y π distribución estacionaria. Considere dos versiones de la cadena en S descrita por P , es decir, tome $X = \{X_n : n = 0, 1, 2, \dots\}$ y $X' = \{X'_n, n = 0, 1, 2, \dots\}$ tales que

- $X = \{X_n : n = 0, 1, 2, \dots\}$ es la versión con matriz de transición P y $\nu(\cdot)$ como distribución inicial, es decir, $P(X_0 = i) = \nu(i), i \in S$.

- $X' = \{X'_n : n = 0, 1, 2, \dots\}$ es la versión con matriz de transición P e inicialmente distribuida de acuerdo con π , es decir, $P(X_0 = i) = \pi(i), i \in S$.
- Considere T un tiempo de paro, $T = \inf\{n : X_n = X'_n\}$ y defina

$$X''_n = \begin{cases} X_n, & \text{si } n < T, \\ X'_n, & \text{si } n \geq T. \end{cases} \quad (2.2)$$

Entonces $\{(X'_n, X''_n) : n \geq 0\}$ es el acoplamiento construido y T es tiempo de acoplamiento. Note que $X_T = X''_T$ para toda $n \geq T$.

Observación: Por definición la distancia de variación total es:

$$d(n) = \max_{A \subset S} |Q_1(A) - Q_2(A)|$$

Se toma solamente $|Q_1(A) - Q_2(A)|$, tomándo los elementos del acoplamiento

$$\begin{aligned} |Q_1(A) - Q_2(A)| &= |P(X_n \in A) - \pi(A)| \\ &= |P(X''_n \in A) - P(X'_n \in A)| \\ &= |P(X''_n \in A, T > n) + P(X''_n \in A, T \leq n) \\ &\quad - P(X'_n \in A, T > n) - P(X'_n \in A, T \leq n)| \\ &\stackrel{(a)}{=} |P(X_n \in A, T > n) - P(X'_n \in A, T > n)| \end{aligned}$$

Si $P(X_n \in A, T > n) - P(X'_n \in A, T > n) > 0$, entonces

$$(1) \dots |P(X_n \in A, T > n) - P(X'_n \in A, T > n)| = P(X_n \in A, T > n) - P(X'_n \in A, T > n)$$

Si $P(X_n \in A, T > n) - P(X'_n \in A, T > n) < 0$, entonces

$$(2) \dots |P(X_n \in A, T > n) - P(X'_n \in A, T > n)| = -P(X_n \in A, T > n) + P(X'_n \in A, T > n)$$

De (1)

$$\begin{aligned} P(X_n \in A, T > n) - P(X'_n \in A, T > n) &\leq P(X_n \in A, T > n) \\ &= P(X_n \in A | T > n) P(T > n) \\ &\leq P(T > n). \end{aligned}$$

De (2) se obtiene bajo el proceso el mismo resultado

$$-P(X_n \in A, T > n) + P(X'_n \in A, T > n) \leq P(T > n).$$

Por lo tanto,

$$|P(X_n \in A) - \pi(A)| \leq P(T > n)$$

tomando el máximo

$$\max_{ACS} |P(X_n \in A) - \pi(A)| \leq \max_{ACS} P(T > n)$$

por lo que se concluye

$$d(n) \leq P(T > n).$$

(a). Para $T \leq n$ sucede que $P(X''_n \in A, T \leq n) = P(X'_n \in A, T \leq n)$ según lo definido en el acoplamiento.

De este modo, por la definición de distancia de variación total, se tiene que $\|P(X_n) - \pi\| \leq P(T > n)$ recuperando así el resultado dado en el corolario 2.4.3

Observación: El tiempo de acoplamiento T es quien otorga la pauta el poder analizar en qué tiempo se da la estacionaridad por medio de la desigualdad de acoplamiento. Entonces cuando T es finito el acoplamiento es considerado un éxito y basta buscar el valor de n^* que hace que la cola de la distribución de T sea "muy pequeña".

2.5 Convergencia exponencial

Considere una cadena de Markov irreducible con espacio de estados S finito, probabilidades de transición P_{ij} , $i, j \in S$ y probabilidades de transición en n pasos $P_{ij}^{(n)}$.

En esta sección se va a demostrar que: para una cadena ergódica, la convergencia al estado estacionario es exponencial. En el capítulo siguiente se estudiará que en ciertos casos la convergencia es en realidad doblemente exponencial. La convergencia exponencial sigue la teoría clásica de Perron-Frobenius donde se tiene que para alguna constante C (no dependiente del

tiempo) y λ el valor propio con mayor valor absoluto de la matriz de transición P , la distancia de variación total $d(n)$ se comporta como $C\lambda^n$ para n suficientemente grande. (Vea Aldous (1983)). Se estudiará ésto con más detalle en el próximo capítulo. El siguiente resultado será utilizado para demostrar la convergencia exponencial.

Proposición 2.5.1. *Si una cadena de Markov es ergódica, con S espacio de estados finito, entonces existen n y $\varepsilon > 0$, tal que $P_{ij}^{(n)} > \varepsilon$, para toda i, j y $n \geq n_0$.*

Demostración. La demostración es inmediata, pues dado que la cadena es ergódica, entonces existen $\pi(i)$, $i \in S$, tales que, $\pi(i) > 0$, para toda $i \in S$ y $\lim_{n \rightarrow \infty} P_{ij}^{(n)} = \pi(j) > 0$. Como el límite es estrictamente positivo, entonces debe existir n_0 , tal que $P_{ij}^{(n)} > \varepsilon > 0$ para toda $i, j \in S$ y $n \geq n_0$. \square

Teorema 2.5.2. *Si $X = \{X_n : n = 0, 1, \dots\}$ es una cadena de Markov ergódica con espacio de estados $S = \{1, 2, \dots, M\}$, matriz de transición P y distribución estacionaria π (def. 1.2.24), entonces*

$$\|P_{ij}^{(n)} - \pi(j)\| \leq (1 - \alpha)^{\frac{n}{M}-1}$$

donde $\alpha = M\varepsilon^2$.

Demostración. Se usará el acoplamiento para demostrar el resultado. De este modo sean $X' = \{X'_n, n = 0, 1, \dots\}$ y $X'' = \{X''_n, n = 0, 1, \dots\}$ dos versiones independientes de la cadena X donde la distribución inicial de X' es

$$\nu_0(x) = P(X'_0 = x) = \begin{cases} 1, & \text{si } x = i \\ 0, & \text{si } x \neq i \end{cases}$$

y la distribución inicial X'' es la distribución estacionaria π , es decir,

$$P(X''_0 = i) = \pi(i), \quad \text{para toda } i \in S.$$

Defina $T = \min\{n : X'_n = X''_n\}$

y $A_k = \{X'_{kN} \neq X''_{kN} | X'_N \neq X''_N, \dots, X'_{(k-1)N} \neq X''_{(k-1)N}\}$

con $k = 1, 2, \dots$ y $A_1 = \{X'_N \neq X''_N\}$.

Si $T > mN$, entonces $X'_N \neq X''_N, X'_{2N} \neq X''_{2N}, \dots, X'_{mN} \neq X''_{mN}$ y se tiene,

$$\begin{aligned} P(T > mN) &\leq P(A_1)P(A_2|A_1)P(A_3|A_2A_1) \cdots P(A_m|A_{m-1} \dots A_1) \\ &= P(X'_N \neq X''_N)P(X'_{2N} \neq X''_{2N}|X'_N \neq X''_N) \cdots \\ &\cdots P(X'_{mN} \neq X''_{mN}|X'_{(m-1)N} \neq X''_{(m-1)N}, \dots, X'_N \neq X''_N) \end{aligned}$$

Por la proposición 2.5.1 se tiene que la probabilidad de que la cadena este en j al tiempo N es de por lo menos ε , no importando el estado presente. Por tanto, dado que las cadenas son independientes la probabilidad de que las dos cadenas se encuentren en j al tiempo N sería ε^2 . La probabilidad de que se encuentren en el mismo estado S es $M\varepsilon^2$ siendo M constante. Por tanto, la probabilidad de no encontrarse en esa situación sería $1 - M\varepsilon^2$.

De este modo

$$P(T > mN) \leq (1 - M\varepsilon^2)^m$$

donde m es un número entero cualquiera. Tome $\alpha = M\varepsilon^2$, de manera que

$$(1 - M\varepsilon^2)^m = (1 - \alpha)^m.$$

Será definida $\bar{X} = \{\bar{X}_n, n \geq 0\}$ cadena de Markov con probabilidad de transición P_{ij} , cuyo estado inicial es elegido de acuerdo a su probabilidad estacionaria, tal que

$$\bar{X} = \begin{cases} X', & \text{si } n < T \\ X'', & \text{si } n \geq T. \end{cases}$$

Note que (\bar{X}, X'') es un acoplamiento y T el tiempo de acoplamiento. De este modo,

$$\|P_{ij}^{(n)} - \pi(j)\| \leq P(T > n) \leq (1 - \alpha)^{\frac{n}{N}-1}$$

Por tanto, X converge exponencialmente al estado estacionario. \square

Capítulo 3

Aplicación a un problema de estructura de datos

"El destino es él que baraja las cartas, pero nosotros somos los que jugamos."

William Shakespeare.

3.1 Introducción

En el capítulo anterior se estudia el uso del acoplamiento para cadenas de Markov, para acotar la distancia entre la distribución de transición en n pasos (o al tiempo t) de una cadena de Markov de la distribución estacionaria. En este capítulo se expondrá la **aplicación del método de acoplamiento en un esquema de auto-organización**, llamado "Movimiento al frente". Este tipo de sistemas aparecen en problemas relacionados con estructuras de datos en el área de cómputo.

Serán presentados el esquema y una forma alternativa de ser visto, bajo el punto de vista de barajeo de cartas. La estrecha relación que guarda este problema con el del "Coleccionador de cupones" se explicará con su debida oportunidad.

Si el lector desea saber más profundamente sobre el tema es recomendable consultar Lindvall (1992).

3.2 Esquema de auto-organización: movimiento al frente

Un esquema de auto-organización se caracteriza por que el mismo dicta su estructura de modo interno, es decir, cualquier cambio sufrido se debe a la forma que interactúa la colección de sus partes. Su estudio describe reglas generales de comportamiento, que ayudan a predecir los cambios en su estructura y a estudiar otros sistemas que exhiben cualidades similares o que pueden ser reducidos a éstas.

La descripción del siguiente esquema de auto-organización mostrará su comportamiento dentro de una de las situaciones donde puede presentarse en realidad; haciendo notar que le hace falta ser preciso en algunos puntos que serán retomados cuando se concrete el modelo:

Suponga que se conoce a un lector que cuenta con su propia y enorme biblioteca. El mantiene ordenados sus libros en estantes de forma que puede reconocer a cada uno verificando sólo el lomo del libro. Mantiene como regla que no lee más de un libro a un tiempo, por lo que hasta que termina de leerlo inicia otro. La selección de temas de lectura no se ve influenciada por ningún parámetro exterior a su propio criterio, ni siquiera por la última lectura realizada. Libro terminado, libro que es inmediatamente puesto en el extremo izquierdo de su librero. Todos los libros tienen la oportunidad de ser elegidos y no comprará ninguno más hasta que haya leído todos los que posee. Cuando decide cual libro va a leer éste es inmediatamente buscado de izquierda a derecha, estante por estante.

Ahora de manera formal:

Suponiendo contar con N objetos, ($N \geq 2$), cada uno de los cuales contará con una etiqueta, por lo que sea definido un conjunto S que represente a estas etiquetas, es decir, $S = \{1, 2, \dots, N\}$. La probabilidad de pedir el objeto con etiqueta i será $p_i > 0$, $i = 1, 2, \dots, N$ con $\sum_{i \in S} p_i = 1$. Es importante mencionar que el caso donde el espacio es equiprobable, es decir, $p_i = \frac{1}{N}$, para todo $i \in S$, será estudiado más adelante.

Considere que estos objetos están almacenados en un estante. Entonces se puede indicar al arreglo de los objetos por un vector N -dimensional

(i_1, i_2, \dots, i_N) , donde $i_k \in S$, $k=1, 2, \dots, N$ e $i_k \neq i_j$ para $j \neq k$.

El conjunto de todos los arreglos posibles es el conjunto S_N , conjunto de todas las diferentes permutaciones¹ de los N elementos de S . Será identificado cada elemento que pertenece a S_N , como el vector de permutación (arreglo) ψ .

Se considera que sólo un pedido es hecho cada vez y siempre después que el pedido anterior es regresado al arreglo. En la medida que los objetos son solicitados y regresados al arreglo se formará una sucesión de arreglos. Cada paso de un arreglo a otro será dictado por el siguiente objeto solicitado, cuando éste regresa será colocado en la posición que se encuentra en el extremo izquierdo del arreglo. Los objetos que se encuentran a la izquierda del objeto demandado son recorridos un sitio a la derecha y los que se hallaban a la derecha se quedan allí. Es claro que cuando el objeto se encuentra en la posición final extrema izquierda no se reacomoda ningún objeto.

La búsqueda de cada demanda es hecha de manera lineal, es decir, si el objeto se encuentra en la posición $2 \leq m \leq N$, las posiciones $1 \leq k \leq m - 1$ son obligatoriamente revisadas.

De esta forma, se identifican dos sucesiones: la de las demandas de objetos y la de la forma en que se disponen estos en el arreglo después de ser modificados por las demandas (esquema de organización). Sean estas sucesiones $R = \{R_n : n = 1, 2, \dots\}$ y $X = \{X_n : n = 0, 1, 2, \dots\}$, sucesión de demandas y sucesión de arreglos respectivamente. Para el caso considerado, la sucesión Markov es una sucesión de demandas independiente, es decir, solicitudes sucesivas son independientes. Las sucesiones R y X tendrán una distribución estacionaria que será indicada por $u(\cdot)$ y $\pi(\cdot)$ y espacio de estados S y S_N respectivamente. Para ver que la sucesión X Markov es estacionaria basta notar que el último objeto pedido se encuentra en la posición más a la izquierda del arreglo y que la probabilidad del próximo arreglo (futuro) sólo depende del arreglo en el presente.

Dado que la solicitud de los objetos es independiente, puede suceder que se solicita el objeto con etiqueta j y en las siguientes ocasiones se repita la petición. En seguida, se dará un ejemplo de como funciona el algoritmo de mover al frente y por lo tanto, como evolucionará R y X .

Ejemplo: Sea $N = 8$ y suponga que originalmente se tiene un arreglo inicial, X_0

$$\text{Arreglo inicial } X_0 = \begin{array}{|c|c|c|c|c|c|c|c|} \hline 1 & 3 & 2 & 5 & 7 & 8 & 4 & 6 \\ \hline \end{array}$$

¹Permutación: número de diferentes formas en que los N objetos pueden ser seleccionados y colocados en N distintas posiciones

$$X_1 = \begin{array}{c} \text{Solicitud: } R_1 = 5 \\ \boxed{5 \mid 1 \mid 3 \mid 2 \mid 7 \mid 8 \mid 4 \mid 6} \end{array}$$

→

$$X_2 = \begin{array}{c} \text{Solicitud: } R_2 = 4 \\ \boxed{4 \mid 5 \mid 1 \mid 3 \mid 2 \mid 7 \mid 8 \mid 6} \end{array}$$

→

$$X_3 = \begin{array}{c} \text{Solicitud: } R_3 = 8 \\ \boxed{8 \mid 4 \mid 5 \mid 1 \mid 3 \mid 2 \mid 7 \mid 6} \end{array}$$

→

$$X_4 = \begin{array}{c} \text{Solicitud: } R_4 = 1 \\ \boxed{1 \mid 8 \mid 4 \mid 5 \mid 3 \mid 2 \mid 7 \mid 6} \end{array}$$

→

$$X_5 = \begin{array}{c} \text{Solicitud: } R_5 = 2 \\ \boxed{2 \mid 1 \mid 8 \mid 4 \mid 5 \mid 3 \mid 7 \mid 6} \end{array}$$

Y así se continua de acuerdo a la sucesión de solicitudes y el esquema de organización.

3.3 Barajeos

Uno de los temas recurrentes en probabilidad es el dedicado a los juegos de azar. En muchas de las ocasiones se entendieron conceptos con ayuda de dados, cartas y fichas de dominó. Ahora, el barajeo funcionará como otra opción para entender el esquema de organización de "Movimiento al frente", dado que este es aplicado a cualquier tipo de objeto.

Un paquete de N cartas es barajado repetidas veces. Dependiendo del barajeo utilizado, esta acción puede ser modelada con una cadena de Markov. Cada vez que se lleva a cabo la mezcla, cada carta tendrá una posición dentro de las N posibles. Es decir, si se introduce un orden en cuanto a posición, la primera puede tener N cartas, en la segunda posición se pueden poner $N - 1$ cartas, en la tercera $N - 2$ cartas, siguiendo así en la última posición queda sólo una carta, de esta forma $N!$ es el número de posibles arreglos que puede tener un mazo de cartas.

posiciones	posibilidades
1	N cartas
2	$N - 1$ cartas
3	$N - 2$ cartas
4	$N - 3$ cartas
	... cartas
N	1 carta

Algunos métodos de barajeo son:

“Top to random”: La carta de la cúspide es tomada y colocada en forma aleatoria dentro del paquete.

“Random to top”: Una carta es elegida de forma aleatoria y puesta en la cúspide.

“Random transposition”: Dos cartas son elegidas aleatoriamente y transpuestas entre sí.

“Top to random, bottom to random”: Se toma la carta de la cúspide, colocándola aleatoriamente en el paquete y después se toma la del final, colocándola también aleatoriamente en cualquier posición del paquete.

“Rifle”: Lo más usual corresponde a cortar el paquete de cartas por la mitad y sus componentes son intercalados entre sí de forma aleatoria.

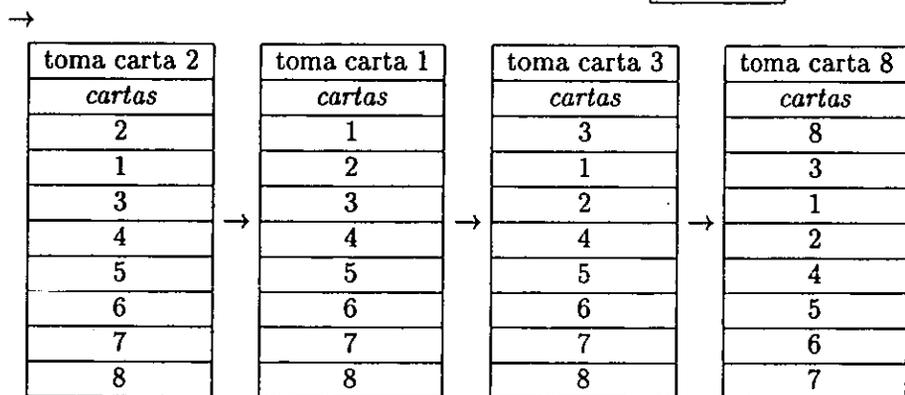
El barajeo “Random to top” a simple vista es el proceso inverso al barajeo “Top to random”, pues según la descripción el tomar una carta de la cúspide e introducirla aleatoriamente es el movimiento inverso a tomar una carta aleatoriamente y colocarla en la cúspide.

El barajeo “Random to top”, cuenta con N cartas etiquetadas $1, 2, \dots, N$. Cada carta es tomada de forma independiente y la carta con etiqueta i tiene la probabilidad $p_i > 0$ de ser seleccionada donde $\sum_{i=1}^N p_i = 1$.

De la descripción del método “Random to top”, se tiene esquemáticamente para $N = 8$, lo siguiente,

Disposición original del mazo de cartas:

<i>etiquetas</i>
1
2
3
4
5
6
7
8



Note que el barajeo es idéntico al esquema de organización de movimiento al frente, pues este puede observarse usando el vector de permutaciones transpuesto.

Las analogías serán el tener una cadena de solicitudes $R = \{R_n : n \geq 1\}$ que en este caso son las cartas tomadas. Una sucesión de arreglos $X = \{X_n : n \geq 0\}$, que es la disposición de las cartas después de cada barajeo, tomando una de las $N!$ permutaciones posibles. Por lo que los espacios de estados para cada una de las sucesiones serán S y S_N respectivamente.

Notas:

- La distribución estacionaria será obtenida en la siguiente sección.
- Con ayuda del lema de distancia de variación total 2.2.3 para procesos reversibles se puede estimar el tiempo n^* , tal que después de n^* barajeos, se tiene el mazo de cartas cerca del estado de equilibrio. Para demostrar los resultados se usará el método de acoplamiento para cadenas de Markov.

- La estimación de n^* puede realizarse utilizando otras técnicas.

3.4 Distribuciones estacionarias

Recordando los elementos a utilizar para construir un acoplamiento, se tiene: las cadenas a acoplar, el tiempo de acoplamiento y las distribuciones estacionarias. Primero se demostrará que las cadenas R y X obtenidas cuando los objetos son tomados independientemente son del tipo ergódico y por tanto, tienen una distribución límite. Se dará una expresión para las distribuciones estacionarias u y π .

Proposición 3.4.1. *Se establece que*

$$u(k) = p_k, k \in S \quad (3.1)$$

es la distribución estacionaria para la sucesión de demandas independientes $R = \{R_n : n = 1, 2, \dots\}$. También se tiene que R es una sucesión Markov reversible en el tiempo.

Demostración. Por la definición de cadena reversible (1.3.1) y por la proposición 1.3.3 basta mostrar que

$$u(k)P_{kj} = u(j)P_{jk}, \text{ para toda } k, j \in S \quad (3.2)$$

y que $\sum_{k \in S} u(k) = 1$,
con $u(k) = p_k$ y P_{ik} las probabilidades de transición de la cadena R .

Dado que R es una sucesión independiente se tiene que $P_{ij} = p_j$ para todo $i, j \in S$. De este modo se tiene que

$$u(k)P_{kj} = p_k p_j = p_j p_k = u(j)P_{jk}$$

$\sum_{k \in S} u(k) = \sum_{k=1}^N p_k = 1$ y el resultado se sigue. □

Teorema 3.4.2. *La distribución estacionaria $\pi(\cdot)$ de la sucesión de arreglos de objetos $X = \{X_n : n = 0, 1, 2, \dots\}$, cuando existe, es igual a*

$$\pi(i_1, i_2, i_3, \dots, i_N) = u(i_1) \prod_{j=2}^N \frac{p_{i_j}}{\sum_{r=j}^N p_{i_r}} \quad (3.3)$$

para $(i_1, i_2, i_3, \dots, i_N) \in S_N$, si y sólo si la sucesión de demandas $R = \{R_n : n = 0, 1, 2, \dots\}$ es una sucesión de Markov con probabilidades de transición $P_{..}$, distribución estacionaria $u(\cdot)$, que satisfacen para $i, k \in S, i \neq k$

$$u(k)(1 - P_{kk})p_i = u(i)(1 - p_k)P_{ik} \quad (3.4)$$

donde $p_i, i = 1, 2, \dots, N, p_i > 0$ y $\sum_{i=1}^N p_i = 1$.

La formulación y prueba de este resultado se encuentra en Rodrigues (1995). A título de ilustración, se hará la demostración del teorema 3.4.2 para el caso $N = 3$.

Demostración: Considere $N = 3$. De esto, se cuentan con tres objetos cualesquiera cuyas etiquetas serán i_1, i_2, i_3 y el conjunto de permutaciones $S_N = \{(i_1, i_2, i_3), (i_1, i_3, i_2), (i_2, i_1, i_3), (i_2, i_3, i_1), (i_3, i_2, i_1), (i_3, i_1, i_2)\}$.

Sea $\psi = (i_1, i_2, i_3)$ y $A = \{(i_1, i_3, i_2), (i_2, i_1, i_3), (i_2, i_3, i_1), (i_3, i_2, i_1), (i_3, i_1, i_2)\}$ y ψ' cualquier elemento en A . Las probabilidades de transición a un paso para llegar al vector ψ serán $P_{\psi\psi} = P_{i_1i_1}$, $P_{\psi'\psi} = P_{i_2i_1}$, con $\psi' \in A$ y 0 en cualquier otro caso.

ψ	demanda(s)
(i_1, i_3, i_2)	i_2, i_1
(i_3, i_1, i_2)	i_2, i_1
(i_3, i_2, i_1)	i_2, i_1
(i_2, i_1, i_3)	i_1
(i_2, i_3, i_1)	i_1
(i_1, i_2, i_3)	i_1

La lectura de esta tabla puede ser seguida como el siguiente ejemplo: Se partirá del vector $\psi' = (i_1, i_3, i_2)$ se demanda i_2 por lo que por el esquema de movimiento al frente se genera el arreglo (i_2, i_1, i_3) , luego entonces se solicita i_1 , resultando $\psi = (i_1, i_2, i_3)$. Entonces como se puede observar con el esquema de movimiento al frente, las probabilidades de transición a un paso son las tres últimas, pues con la demanda del objeto de etiqueta i_1 en seguida, se consigue pasar del vector ψ a ψ y del vector ψ' a ψ .

El método usado en la demostración es el siguiente: se supone que existe $\pi(\cdot)$ y por tanto satisface la ecuación de balance total

$$\pi(\psi) = P_{\psi\psi}\pi(\psi) + \sum_{\psi' \in A} P_{\psi'\psi}\pi(\psi') \quad (3.5)$$

$$\sum_{\psi \in S_N} \pi(\psi) = 1 \quad (3.6)$$

De la ecuación (3.5) se genera,

$$\pi(\psi)(1 - P_{\psi\psi}) = \sum_{\psi' \in A} P_{\psi'\psi} \pi(\psi')$$

y sustituyendo los valores de las probabilidades de transición, se sigue

$$\begin{aligned} \pi(\psi)(1 - P_{i_1 i_1}) &= P_{i_2 i_1} \sum_{\psi' \in A} \pi(\psi') \\ \pi(i_1, i_2, i_3)(1 - P_{i_1 i_1}) &= P_{i_2 i_1} \left[\pi(i_2, i_1, i_3) + \pi(i_2, i_3, i_1) \right]. \end{aligned}$$

Atendiendo al hecho de la existencia de $\pi(\cdot)$ y tomando en cuenta que es de la forma dada en la ecuación anterior, se tiene que

$$\begin{aligned} (1 - P_{i_1, i_1})\pi(i_1, i_2, i_3) &= P_{i_2, i_1} \left[\pi(i_2, i_1, i_3) + \pi(i_2, i_3, i_1) \right] \\ &= P_{i_2 i_1} \left[u(i_2) \times \frac{p_{i_1}}{p_{i_1} + p_{i_3}} + u(i_2) \times \frac{p_{i_3}}{p_{i_1} + p_{i_3}} \right] \\ &= \frac{P_{i_2 i_1} u(i_2)}{p_{i_1} + p_{i_3}} (p_{i_1} + p_{i_3}) \\ &= P_{i_2 i_1} u(i_2) \end{aligned}$$

Entonces

$$(1 - P_{i_1, i_1})\pi(i_1, i_2, i_3) = P_{i_2 i_1} u(i_2)$$

y por tanto

$$\begin{aligned} \pi(i_1, i_2, i_3) &= \frac{P_{i_2 i_1} u(i_2)}{1 - P_{i_1 i_1}} \\ &= \frac{P_{i_2 i_1} u(i_2)}{1 - P_{i_1 i_1}} \times \frac{p_{i_3}}{p_{i_3}} \end{aligned}$$

De este modo

$$\pi(i_1, i_2, i_3) = u(i_1) \frac{p_{i_2}}{1 - p_{i_1}} \times \frac{p_{i_3}}{1 - p_{i_1} - p_{i_2}},$$

si y solo si

$$\frac{P_{i_2 i_1} u(i_2)}{1 - P_{i_1 i_1}} = \frac{p_{i_2} u(i_1)}{1 - p_{i_1}},$$

o equivalentemente si y solo si

$$P_{i_2 i_1} u(i_2)(1 - p_{i_1}) = u(i_1) p_{i_2} (1 - P_{i_1 i_1})$$

que corresponde a la condición del teorema 3.4.2

$$P_{ik} u(i)(1 - p_k) = u(k) p_i (1 - P_{kk}),$$

y que es la condición que pide el teorema. Además, falta $\sum_{\psi \in S_N} \pi(\psi) = 1$ con $N = 3$.

$$\begin{aligned} \sum_{\psi \in S_N} \pi(\psi) &= \pi(i_1, i_2, i_3) + \pi(i_1, i_3, i_2) + \pi(i_2, i_1, i_3) + \pi(i_2, i_3, i_1) \\ &\quad + \pi(i_3, i_2, i_1) + \pi(i_3, i_1, i_2). \end{aligned}$$

Sustituyendo la expresión de π se tiene

$$\begin{aligned} \sum_{\psi \in S_N} \pi(\psi) &= u(i_1) \left[\frac{p_{i_2}}{p_{i_2} + p_{i_3}} \times \frac{p_{i_3}}{p_{i_3}} + \frac{p_{i_3}}{p_{i_3} + p_{i_2}} \times \frac{p_{i_2}}{p_{i_2}} \right] \\ &\quad + u(i_2) \left[\frac{p_{i_1}}{p_{i_1} + p_{i_3}} \times \frac{p_{i_3}}{p_{i_3}} + \frac{p_{i_3}}{p_{i_3} + p_{i_1}} \times \frac{p_{i_1}}{p_{i_1}} \right] \\ &\quad + u(i_3) \left[\frac{p_{i_1}}{p_{i_1} + p_{i_2}} \times \frac{p_{i_2}}{p_{i_2}} + \frac{p_{i_2}}{p_{i_2} + p_{i_1}} \times \frac{p_{i_1}}{p_{i_1}} \right] \end{aligned}$$

factorizando y agrupando, se sigue

$$\sum_{\psi \in S_N} \pi(\psi) = \sum_{k \in S} u(k) = 1$$

dado que $u(\cdot)$ es la distribución estacionaria de R .

Corolario 3.4.3. *La distribución estacionaria $\pi(\cdot)$ de la sucesión de arreglos de objetos $X = \{X_n : n = 0, 1, 2, \dots\}$ con R una sucesión independiente, existe y es dada por*

$$\pi(i_1, i_2, \dots, i_N) = p_{i_1} \prod_{j=2}^N \frac{p_{i_j}}{\sum_{r=j}^N p_{i_r}} \quad (3.7)$$

para todo $(i_1, i_2, \dots, i_N) \in S_N$.

Demostración. Para R una sucesión independiente se tiene que $P_{ik} = p_k, i, k \in S$ y por la proposición 3.4.1 se tiene que $u_k = p_k, k \in S$. De este modo se ha observado que la condición del teorema 3.4.2 es satisfecha dado que para R una sucesión de demandas independientes se tiene que

$$u(k)(1 - P_{kk}) = p_k(1 - p_k)p_i = p_i(1 - p_k)p_k = u(i)(1 - p_k)P_{ik}. \quad (3.8)$$

□

Observación: Cuando el espacio es equiprobable, las probabilidades de transición son definidas por:

$$P_{ij} = p_j = \frac{1}{N} \quad (3.9)$$

para toda $i, j \in S$. Puesto que la posición que guarda cada objeto no depende del sitio del cual proviene; luego, la probabilidad de estar en una posición j debe de cumplir que $p_i = p_j$, para toda $i, j \in S$, $\sum_{j \in S} p_j$, por lo cual, $p_j = \frac{1}{N}$, donde N es el número total de estados de S .

$$\begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \frac{1}{N} & \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \frac{1}{N} & \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N} & \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{pmatrix}$$

Con estas probabilidades de transición, se ha definido una matriz de transición que es ergódica dado que es aperiódica (pues $P_{ii} = \frac{1}{N}$ para todo $i \in S$), irreducible ($P_{ij} \neq 0$ para todo $i, j \in S$) y tiene espacio de estados finito (entonces es recurrente y por la forma de la matriz de transición se nota que es recurrente positiva). Así se tiene que la cadena X es irreducible y aperiódica con espacio de estados finito. De este modo todos sus estados son recurrentes positivos. Por tanto, la sucesión de arreglos es ergódica y la distribución límite existe.

3.5 Acoplamiento

En la sección anterior, se ha establecido las dos distribuciones estacionarias para nuestras cadenas. Ahora, se va a aplicar el método de acoplamiento para estimar el tiempo que tarda la cadena X en llegar al estado de equilibrio.

Se sabe que (ver Aldous 1983) para dos versiones independientes de una misma cadena de Markov ergódica, que son puestas a evolucionar independientemente existe un tiempo T^* tal que ellas se encuentran en el mismo estado. Pero lo interesante es hacerlas evolucionar conjuntamente de modo que se pueda analizar el tiempo de acoplamiento T . Esto se debe a que, al acotar la distancia de variación total por la cola de la distribución del tiempo de acoplamiento, se necesita tener información suficiente con el fin de estudiar el comportamiento de esta distribución, para poder estimar el tiempo n^* tal que $P(T > n^*)$ sea "muy pequeña" para T grande.

En esta sección, se va definir el **acoplamiento** que se utiliza durante lo restante de este capítulo.

Definición 3.5.1. i. Sean $X = \{X_n : n = 0, 1, 2, \dots\}$ y $X' = \{X'_n : n = 0, 1, 2, \dots\}$ dos versiones de la sucesión de permutaciones.

ii. Sean $\pi(\cdot)$ y $P_{i_0}(X_n = \cdot)$ sus distribuciones iniciales, respectivas, donde $\pi(\cdot)$ es la distribución estacionaria.

iii. Sea $R = \{R_n : n \geq 0\}$, la sucesión de demandas.

iv. Sea T definido por $T = \min\{n : \{R_1, \dots, R_n\} = S\}$. Este será el tiempo de acoplamiento, estudiando la razón adelante.

El acoplamiento será construido entre las dos sucesiones de permutaciones que comparten el espacio de estados S_N pues son quienes sufren el cambio directo cuando se le aplica el esquema de organización ó barajeo. Además son ellas quienes descubren la cantidad de tiempo que tarda para llegar al equilibrio. Note que el tiempo de acoplamiento se basa en las solicitudes $\{R_n : n \geq 0\}$, de todos los diferentes objetos que forman S . De hecho, si se busca una igualdad entre las cadenas X y X' en este tiempo, el lograrla depende de involucrar a todo objeto del espacio de estados en la tarea de ser acomodado y ser regresado utilizando el algoritmo de movimiento al frente. Se hace la observación de que T no es el tiempo óptimo, pues puede lograrse la igualdad entre arreglos con mas prontitud, pero el tiempo T definido aquí es posible de analizar.

En sí el porque funciona este acoplamiento queda explicado por medio de la aplicación del esquema de movimiento hacia al frente en conjunto por la aplicación de la sucesión de demandas (o la visión alternativa del barajeo). Pues no importando como se inician las sucesiones $X = \{X_n : n = 0, 1, 2, \dots\}$ y $X' = \{X'_n : n = 0, 1, 2, \dots\}$, eventualmente se tendrá el mismo arreglo.

Se explicará el acoplamiento que se usará primero por medio de un ejemplo y después formalmente. De este modo con $N = 4$, iniciando con dos distintos arreglos y aplicando lo ya descrito.

Primer arreglo original $X_0 =$

solicitud	pos.1	pos.2	pos.3	pos.4
	1	2	4	3

Segundo arreglo original $X'_0 =$

solicitud	pos.1	pos.2	pos.3	pos.4
	2	4	1	3

Primer arreglo en la primera solicitud

solicitud	pos.1	pos.2	pos.3	pos.4
3	3	1	2	4

Segundo arreglo en la primera solicitud

solicitud	pos.1	pos.2	pos.3	pos.4
3	3	2	4	1

Primer arreglo en la segunda solicitud

solicitud	pos.1	pos.2	pos.3	pos.4
1	1	3	2	4

Segundo arreglo en la segunda solicitud

solicitud	pos.1	pos.2	pos.3	pos.4
1	1	3	2	4

Primer arreglo en la tercera solicitud

solicitud	pos.1	pos.2	pos.3	pos.4
2	2	1	3	4

Segundo arreglo en la tercera solicitud

solicitud	pos.1	pos.2	pos.3	pos.4
2	2	1	3	4

Primer arreglo en la cuarta solicitud

solicitud	pos.1	pos.2	pos.3	pos.4
4	4	2	1	3

Segundo arreglo en la cuarta solicitud

solicitud	pos.1	pos.2	pos.3	pos.4
4	4	2	1	3

Ahora, se iniciará la descripción del acoplamiento a usar. Tome las dos cadenas X y X' y aplique a ellas la misma sucesión de demandas R . Note que después de que el primer pedido es regresado al arreglo, los dos arreglos coincidirían en la primera coordenada (pues es en esta posición que el objeto es regresado y dado que se usa la misma demanda para ambos arreglos, el objeto que esta en la primera posición es el mismo en ambos arreglos). Después que la segunda demanda es regresada al arreglo tendrá que X y X' coinciden en las dos posiciones de la izquierda (si el objeto demandado es diferente del primero) o coinciden solamente en la primera posición (si el objeto demandado es el mismo que el primero). Prosiguiendo de este modo, es decir, usando la misma sucesión de demandas para X y X' , hasta que todos los objetos hayan sido demandados por lo menos una vez. Note que después que se demanda cada objeto por lo menos una vez, se tiene que el arreglo X es el mismo que X' y proseguirá de este modo. Así, que el tiempo T , definido en (ii) arriba, es el tiempo de acoplamiento. (Ver Aldous (1983) y Rodrigues (1995))

3.6 El coleccionador de cupones

Se enunciará, ahora, el problema del coleccionador de cupones.

En su momento, se verá cómo este problema se relaciona con el tiempo de acoplamiento T . Existen varias versiones de este problema y en Boneh y Hofri se puede encontrar una lista de aplicaciones de este problema incluyendo un ejemplo en lingüística. La descripción hecha aquí es la más sencilla.

Se tiene una caja con N pelotas, éstas se encontrarán etiquetadas de 1 hasta N . El evento consiste en meter la mano a la caja, sacar una pelota, anotar la etiqueta en una lista y regresar la pelota a la caja. De este modo se tiene un muestreo con reemplazo; considerando un éxito cada vez que se saca (y se anota) una pelota que no ha sido elegida anteriormente. ¿cuál es el número de veces (V) que se mete la mano hasta que cada una de las pelotas haya sido tomada (anotada) al menos una vez?

Para el presente estudio, se va a considerar el caso donde las pelotas son muestreadas independientemente y tienen distribución uniforme, es decir, una pelota tiene probabilidad $\frac{1}{N}$ de ser retirada, independientemente de su etiqueta. Esta suposición será considerada el resto de este capítulo.

Cuando se juegue ocurrirá que las probabilidades de tener éxito estarán

definidas por:

**ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA**

- la primera vez que metes la mano obtendrás el primer éxito $\frac{N}{N}$, dado que de N posibilidades puedes elegir N objetos aún no demandados.
- el segundo éxito sucederá con probabilidad $\frac{N-1}{N}$, pues se tienen $N - 1$ pelotas distintas a las anotadas entre las N .
- Se prosigue de este modo, teniendo que el k -ésimo éxito sucederá con probabilidad $\frac{N-k}{N}$ y
- el N -ésimo éxito sucederá con probabilidad $\frac{1}{N}$, pues se tiene solamente una pelota diferente de las anotadas a la lista entre las N pelotas.

Sea definido lo siguiente $V_i, i = 0, \dots, N - 1$, la variable aleatoria que representa el número de total de ocasiones en las que se debe realizar el evento para que se pase de tener i anotaciones a $i + 1$ anotaciones con la probabilidad de éxito ya descrita. De este modo, $V_0 = 1$ pues como no se ha extraído alguna bola, la primera extracción indica la primera anotación. V_1 es el número de extracciones que se dan de la primera anotación para obtener la segunda anotación. Se debe recordar que las pelotas anotadas son extracciones que no se han hecho anteriormente.

Como se había acordado, V es el número total de ocasiones que realizas el evento para obtener N distintas anotaciones. De este modo se tiene que se puede escribir V del siguiente modo.

$$V = V_0 + V_1 + V_2 + \dots + V_{N-1}$$

Por tanto, el número esperado de ocasiones a realizar el evento es:

$$\begin{aligned}
 E[V] &= E[v_0 + v_1 + v_2 + \dots + v_{N-1}] \\
 &= \sum_{i=0}^{N-1} E(v_i) \\
 &= \sum_{i=0}^{N-1} \frac{1}{\frac{N-i}{N}} \\
 &= \sum_{i=0}^{N-1} \frac{N}{N-i} \\
 &= N \sum_{i=0}^{N-1} \frac{1}{N-i}
 \end{aligned}$$

para N muy grande se puede aproximar $\sum_{i=1}^{N-1} \frac{1}{N-i}$ por $\ln(N)$, entonces

$$E[V] \doteq N \ln N + O(N).$$

donde $X \doteq$ significa que X es aproximadamente y $O(N)$ es una función tal que

$$\lim_{N \rightarrow \infty} \frac{O(N)}{N} = c$$

donde c es una constante.

3.7 Resultados Principales

En este apartado se reúnen algunos los resultados que son logrados basándose en la técnica de acoplamiento para estimar el tiempo de corte para la cadena de Markov $X = \{X_n : n = 0, 1, 2, \dots\}$ que registra los arreglos de los objetos. Se asume que las demandas son independientes y que $p_i = \frac{1}{N}$, $i \in \{1, 2, \dots, N\} = S$. Los resultados son útiles para una N muy grande y son los siguientes.

Lema 3.7.1. Si la sucesión de solicitudes $R = \{R_n : n = 0, 1, 2, \dots\}$ es independiente y el espacio de objetos solicitados es equiprobable (es decir, $p_i = \frac{1}{N}, i \in \{1, 2, \dots, N\} = S$), entonces el tiempo de acoplamiento T es tal que

T se distribuye como V ,

donde V es el tiempo hasta que cada uno de los objetos haya sido solicitado al menos una vez en el problema del coleccionador de cupones con N objetos de la urna.

Demostración. Sea $R = \{R_n : n = 0, 1, 2, \dots\}$ la sucesión de solicitudes. Defina $S_i, i = 0, 1, 2, \dots$ serán los tiempos aleatorios, tales que registran los tiempos en los cuales objetos no previamente solicitados son demandados. De este modo,

$$S_i = \begin{cases} 0, & \text{para } i = 0 \\ 1, & \text{para } i = 1 \\ \inf\{j : j > S_{i-1}, R_j \notin \{R_1, \dots, R_{j-1}\}\}, & \text{para } i = 2, \dots, N \end{cases} \quad (3.10)$$

Será definida $T_i = S_i - S_{i-1}$, con $i = 1, \dots, N$, como la variable que registra el número de solicitudes hechas entre la demanda de $i - 1$ e i objetos no previamente solicitados. Bajo esta definición se tiene que $P(T_1 = S_1 - S_0 = 1) = 1$, para $i = 2, \dots, N$.

De forma general, la probabilidad de que el número de solicitudes entre el tiempo que se tiene $i - 1$ objetos registrados e i . Para $T_i = j$, se debe tomar en las primeras $(j - 1)$ extracciones alguno de los $i - 1$ objetos que hayan sido previamente seleccionados y en la j -ésima extracción se debe tomar uno de los objetos $N - (i - 1)$ que no han sido demandados. Dado que se tienen objetos uniformemente seleccionados, se tiene que:

$$P(T_i = j) = \left(1 - \frac{N - (i - 1)}{N}\right)^{(j-1)} \left(\frac{N - (i - 1)}{N}\right)$$

con $j = 1, 2, \dots$

Entonces el tiempo de acoplamiento que es definido como el tiempo de espera hasta haber solicitado todos y cada uno de los objetos, al menos una vez, puede ser reexpresado en los siguientes términos:

$$T = T_1 + T_2 + T_3 + \dots + T_N$$

donde, T_i se distribuye como U_{i-1} , con $i = 1, 2, \dots, N$. U_{i-1} es el número de eventos requeridos entre las ocasiones donde de $i - 1$ objetos e i objetos no previamente elegidos son tomados dentro del problema del coleccionador de cupones. \square

A seguir serán presentados algunos lemas que serán utilizados en la demostración de los teoremas relacionados con la tasa de convergencia al estado estacionario para cadena de arreglos X .

Dentro del problema del coleccionador de cupones, sea $W_N(m)$ el tiempo de espera para que por primera vez se tengan anotadas $N - m$ distintas pelotas de un total de N pelotas, con $N > m$ y $m \geq 0$.

Entonces $W_N(0)$ es el tiempo de espera para que se obtengan N pelotas anotadas y donde su total es de N pelotas distintas.

Lema 3.7.2. (*Csörgö, (1993)*)

$$\lim_{N \rightarrow \infty} P(W_N(0) \leq N \ln N + Nx) = G_0(x) = e^{-e^{-x}}$$

con $-\infty < x < \infty$.

Lema 3.7.3. *Sea V , el número de repeticiones requeridas para que cada una de las pelotas sea tomada, al menos en una ocasión, entonces*

$$P(V > N \ln N + cN) \leq e^{-c}$$

para $c \geq 0, N \geq 1$.

Demostración. Sea $m = N \ln N + cN$. Para todo objeto $i = 1, 2, \dots, N$, sea

A_i : objeto rotulado i no elegido dentro de las primeras m demandas.

Entonces

$$\begin{aligned} P(V > m) &= P\left(\bigcup_{i=1}^N A_i\right) \\ &\leq \sum_{i=1}^N P(A_i) \\ &= N \left(1 - \frac{1}{N}\right)^m \end{aligned}$$

Analizando el comportamiento de $(1 - \frac{1}{N})^m$ por separado se tiene lo siguiente,

$$\begin{aligned}
 \left(1 - \frac{1}{N}\right)^m &\stackrel{(1)}{=} \sum_{i=0}^m \left(\frac{-1}{N}\right)^i \binom{m}{i} \\
 &= \sum_{i=0}^m \left(\frac{-1}{N}\right)^i \frac{m!}{(m-i)!i!} \\
 &= \sum_{i=0}^m \left(\frac{-1}{N}\right)^i \frac{m(m-1)\cdots(m-i+1)(m-i)!}{(m-i)!i!} \\
 &= \sum_{i=0}^m \left(\frac{-1}{N}\right)^i \frac{m(m-1)\cdots(m-i+1)}{i!} \\
 &\stackrel{(2)}{\leq} \sum_{i=0}^N \left(\frac{-1}{N}\right)^i \frac{m^i}{i!} \\
 &\leq \sum_{i=0}^N \left(\frac{-m}{N}\right)^i \frac{1}{i!} \\
 &\leq \sum_{i=0}^{\infty} \left(\frac{-m}{N}\right)^i \frac{1}{i!}
 \end{aligned}$$

(1). Expansión binomial.

(2). $m(m-1)\cdots(m-i+1) \leq m \cdot m \cdots m = m^i$.

Ahora, note que $\sum_{i=0}^{\infty} \left(\frac{-m}{N}\right)^i \frac{1}{i!} = e^{\left(\frac{-m}{N}\right)}$.

Por tanto,

$$\left(1 - \frac{1}{N}\right)^m \leq e^{\left(\frac{-m}{N}\right)}$$

entonces

$$N\left(1 - \frac{1}{N}\right)^m \leq Ne^{\left(\frac{-m}{N}\right)}.$$

Tomando $m = N \ln N + cN$, donde $c \geq 0$ es una constante. En relación a N , se tiene

$$\frac{-m}{N} = \frac{-N \ln N - cN}{N} = \ln N^{-1} - c$$

y por tanto,

$$N \left(1 - \frac{1}{N}\right)^{N \ln N + cN} \leq e^{-c}$$

de donde sigue que

$$P(V > N \ln N + cN) \leq N \left(1 - \frac{1}{N}\right)^{(N \ln N + cN)} \leq e^{-c}$$

como se quería demostrar. \square

Teorema 3.7.4. *Si la sucesión de solicitudes $R = \{R_n : n = 0, 1, 2, \dots\}$ es independiente y los objetos son solicitados uniformemente, es decir, $p_i = \frac{1}{N}$, $i = 1, 2, \dots$, entonces la distancia de variación total es tal que*

1. $d(N \ln N + cN) \leq e^{-c}$ para $c > 0$.
2. $\limsup_{N \rightarrow \infty} d(N \ln N + cN) \leq 1 - e^{-e^{-c}}$, para toda c .
3. Si $c_N \rightarrow \infty$ cuando $N \rightarrow \infty$, entonces $\lim_{N \rightarrow \infty} d(N \ln N - c_N N) = 1$.

Demostración. Sean $m = N \ln N + cN$ y $m' = N \ln N - c_N N$, entonces

1. Por la desigualdad de acoplamiento se tiene que para T el tiempo de acoplamiento

$$d(n) \leq P(T > n)$$

para toda n , entonces es válido para $n = m$. Por el lema 3.7.1, T se distribuye como V del problema del coleccionador de cupones y por el lema 3.7.3 se sigue que

$$P(T > m) = P(V > m) \leq e^{-c}$$

y de este modo se concluye que

$$d(N \ln N + cN) \leq e^{-c}.$$

2. Por el lema 3.7.2 se sabe que para el problema del coleccionador de cupones

$$\lim_{N \rightarrow \infty} P(W_N(0) \leq N \ln N + cN) = e^{-e^{-c}}.$$

Por la definición de $W_N(0)$ y V (Lemas 3.7.2 y 3.7.3) se asume, que son lo mismo, de donde, se concluye que

$$\lim_{N \rightarrow \infty} P(V \leq N \ln N + cN) = e^{-e^{-c}}$$

por tanto,

$$\begin{aligned} \lim_{N \rightarrow \infty} P(V > N \ln N + cN) &= 1 - \lim_{N \rightarrow \infty} P(V \leq N \ln N + cN) \\ &= 1 - e^{-e^{-c}}. \end{aligned}$$

Por la desigualdad de acoplamiento (corolario 2.4.3) y del lema 3.7.2 sigue que

$$d(m) \leq P(V > m).$$

Por tanto,

$$\limsup_{n \rightarrow \infty} d(N \ln N + cN) \leq 1 - e^{-e^{-c}}$$

como se quería demostrar.

3. Considere ahora m' con $c_N \rightarrow \infty$ cuando $N \rightarrow \infty$. Se fija un j arbitrario en $S = \{1, 2, \dots, N\}$ y sea $B_j \subseteq S_N$ quien señale al conjunto de todas las permutaciones (arreglos) que conserven el orden relativo original de los j objetos que se cuentan en el extremo derecho de la permutación (arreglo).

Una de las formas de encontrar el límite inferior para la distancia de variación total consiste en tomar un conjunto B adecuado y por la definición de distancia de variación total 2.2.1 se tiene

$$d(n) \geq |P(X_n \in B) - \pi(B)| \geq P(X_n \in B) - \pi(B)$$

Sea $B = B_j$ definido arriba. Se tienen $j!$ formas de preservar el orden relativo de esos j objetos y como los arreglos son equiprobables (corolario 3.4.3), entonces $\pi(B_j) = \frac{1}{j!}$, de donde

$$d(n) \geq P(X_n \in B_j) - \frac{1}{j!} \quad (3.11)$$

Sea $S_{N-j+1}^{(N)}$ el tiempo, dentro del problema del coleccionador de cupones con N cupones, donde $N - j + 1$ cupones han sido elegidos al menos una vez. Por tanto, $j - 1$ cupones no han sido elegidos, por lo que conservan su orden relativo.

Por definición de B_j , se puede deducir que en este conjunto se incluye a cierto número de arreglos donde $j - 1$ conservan su orden relativo en el extremo derecho, entonces se tiene la siguiente relación entre eventos

$$\{S_{N-j+1}^{(N)} > m'\} \subseteq \{X_{m'} \in B_j\}$$

y por tanto,

$$P(X_{m'} \in B_j) \geq P(S_{N-j+1}^{(N)} > m') \quad (3.12)$$

Es útil recordar los conceptos definidos para T_i y $\{S_i, i = 0, 1, 2, \dots\}$ en el lema 3.7.1. Se sabe que $T_i = S_i - S_{i-1}, i = 1, 2, \dots, N$, entonces se reescribe $S_n = \sum_{i=1}^n T_i$, ocurriendo que

$$E[S_{N-j+1}^{(N)}] = \sum_{i=1}^{N-j+1} E[T_i] = N \ln N + O(N)$$

$$\text{Var}[S_{N-j+1}^{(N)}] = \sum_{i=1}^{N-j+1} \text{Var}[T_i] = O(N^2).$$

Note que

$$P(S_{N-j+1}^{(N)} \leq N \ln N - Nc_N) \leq P(S_{N-j+1}^{(N)} \leq N \ln N + O(N) - Nc_N).$$

Por tanto,

$$P(S_{N-j+1}^{(N)} - N \ln N - O(N) \leq -Nc_N) \quad (3.13)$$

$$\leq P(S_{N-j+1}^{(N)} - (N \ln N) - O(N) \leq -Nc_N)$$

$$+ P(S_{N-j+1}^{(N)} - (N \ln N) - O(N) \geq Nc_N)$$

$$\stackrel{(1)}{=} P(|S_{N-j+1}^{(N)} - (N \ln N) - O(N)| \geq Nc_N)$$

$$= P(|S_{N-j+1}^{(N)} - E[S_{N-j+1}^{(N)}]| \geq Nc_N)$$

$$\stackrel{(2)}{\leq} \frac{\text{Var}(S_{N-j+1}^{(N)})}{N^2 C_N^2}$$

(1). Por definición $|x| \geq r$ si y solo si $x \leq -r$ ó $x \geq r$.

(2). Usando la desigualdad de Chebyshev.

Note que

$$\lim_{N \rightarrow \infty} \frac{\text{Var}(S_{N-j+1}^{(N)})}{N^2 C_N^2} = \frac{O(N^2)}{N^2 C_N^2} = 0$$

y por tanto, tomando el límite superior en la ecuación (3.13) se tiene que

$$\lim_{N \rightarrow \infty} \sup P(S_{N-j+1}^{(N)} - N \ln N - O(N) \leq -N c_N) \leq 0$$

Dado que una función de distribución de probabilidad se encuentra entre 0 y 1, sigue que

$$\lim_{N \rightarrow \infty} P(S_{N-j+1}^{(N)} \leq N \ln N - N c_N + O(N)) = 0,$$

de donde, se tiene que

$$\begin{aligned} \lim_{N \rightarrow \infty} P(S_{N-j+1}^{(N)} > N \ln N - N c_N + O(N)) \\ \geq 1 - \lim_{N \rightarrow \infty} P(S_{N-j+1}^{(N)} \leq N \ln N - N c_N + O(N)) = 1. \end{aligned}$$

Por la ecuación (3.12) sigue que:

$$P(X_{m'} \in B_j) \geq P(S_{N-j+1}^{(N)} > m') = 1$$

además por la ecuación (3.11) se tiene que,

$$P(X_{m'} \in B_j) - \frac{1}{j!} \leq d(m').$$

Tomando j suficientemente grande, tal que $\frac{1}{j!}$ se aproxime a 0, sigue que

$$1 \leq P(X_{N \ln N - N c_N + O(N)} \in B_j) - \frac{1}{j!} \leq d(N \ln N - N c_N)$$

Entonces $\lim_{N \rightarrow \infty} \inf d(N \ln N - Nc_N) = 1$ y como $d(m') \in [0, 1]$

$$\lim_{N \rightarrow \infty} d(N \ln N - Nc_N) = 1$$

y el teorema está demostrado. □

Observación: Note que el teorema 3.7.4 dice que, para $m = N \log N + cN$, $c > 0$, la distancia entre la probabilidad de transición al tiempo m y la distribución estacionaria decrece con una rapidez doblemente exponencial. De este modo, se tiene que $n^* = N \ln N$ en el tiempo de corte para la cadena de arreglos X . Es decir, para $n = n^* + \epsilon$ se tiene que la distancia de variación total es muy cercana a cero y para $n = n^* - \epsilon$ la distancia de variación total es muy cercana a uno.

Capítulo 4

Comentarios

“He who has gone furthest has a long way to go.”

Robin Fulton

4.1 Introducción

Este capítulo presenta como fin el informar de detalles que se juzgan importantes, de modo que la mayoría de lo que será presentado no contará con una demostración formal, aunque en algunos casos se dará un breve esbozo de su realización.

Los comentarios serán de carácter concluyente sobre lo desarrollado en esta tesis, mostrando alguna otra postura ó técnica en cuanto al mismo tratamiento del tiempo de corte para cadenas de Markov o mostrando un tópico relacionado.

4.2 Probabilidades no uniformes

Otra forma de cambio en las condiciones del problema de barajeo o colección de cupones o el esquema de auto-organización es el trabajar con probabilidades de solicitud no uniforme, es decir, tomar las probabilidades de selección $p_i \neq \frac{1}{N}$ para algún o todo $i \in \{1, 2, \dots, N\}$. De este modo, se fija

el mecanismo de pedidos (independencia) y el sistema auto-organizador y se cambian las probabilidades de las solicitudes.

En esta orientación, el artículo de Fill (1996), establece teoremas y comentarios valiosos. El utiliza otras probabilidades ligado a la aproximación al tiempo de corte que otorga la distancia de separación total y la distancia de variación total. Así, sólo se enuncia el teorema y se exhorta a quien desee mayor información a la lectura de la referencia.

Se considera que el sistema organizador utilizado es el movimiento al frente y que los objetos son solicitados independientemente.

Teorema 4.2.1. (i) Ley de Zipf: Si $p_i = (iH_N)^{-1}$, con $H_N = \sum_{i=1}^N \frac{1}{i}$, entonces el tiempo aproximado de corte es $N \log N (\log N + c)$.

(ii) Decaimiento Geométrico: Sea $0 < q = 1 - p < 1$. Si $p_i = pq^{-1}$ para $i = 1, \dots, N-1$ y $w_N = q^{N-1}$, entonces el tiempo de corte es cq^{-N} .

En el caso de un p_j general, Aldous (1989), establece un planteamiento para probabilidades desiguales, donde cada objeto j tiene una probabilidad p_j . El establece que el tiempo aproximado de corte será la solución t de

$$\sum_{i \in S} e^{-p_i t} = 1$$

La obtención de esa solución es el problema, pues en algunas ocasiones la expresión se torna compleja.

Un caso sencillo es el siguiente: se tiene $S = \{1, 2, \dots, N\}$, con $p_i = \frac{1}{N}$, $i \in S$, es decir, se tiene p_i la distribución uniforme en S . Se ve cual es el tiempo t^* que satisface

$$\begin{aligned} \sum_{i \in S} e^{(-\frac{1}{N}t^*)} &= 1 \\ N e^{(-\frac{1}{N}t^*)} &= 1 \end{aligned}$$

ó

$$e^{-\frac{1}{N}t^*} = \frac{1}{N}$$

Aplicando la función logarítmica se tiene

$$-\frac{1}{N}t^* = \ln \frac{1}{N} = \ln 1 - \ln N$$

$$-\frac{1}{N}t^* = -\ln N$$

y por tanto

$$t^* = N \ln N$$

De este modo, se observa que el tiempo de corte t^* es aproximadamente $N \ln N$ y esto es lo que establece el problema del coleccionador de cupones.

4.3 Sucesiones de demandas que son una sucesión de Markov

El esquema de organización de movimiento al frente analizado como un ejemplo de aplicación del tiempo de acoplamiento, no tuvo ningún tipo de restricción en las solicitudes, éstas eran independientes. Existen y han sido analizados otros casos, estos otros casos cambian la forma en que hacen las solicitudes. Por ejemplo, en Rodrigues (1995) se tienen las siguientes modificaciones:

(1). Durante un número k dado de demandas no se permite la solicitud a un objeto, que ya apareció en una de las siguientes demandas previas. A partir de este hecho se establecen las probabilidades de transición y las distribuciones estacionarias tanto para la sucesión de demandas como para la de arreglos.

(2). Otra forma de modificación es el tipo de solicitud es la de correlación $-p$ que implica un tipo de dependencia entre las sucesivas demandas de la siguiente forma: si un objeto es seleccionado en el presente, entonces con probabilidad $p \in (0, 1)$ el mismo objeto es seleccionado en la próxima demanda. Caso contrario, es decir, con probabilidad $(1 - p)$ una demanda independiente es hecha.

4.4 Tiempo conocido y tiempo analizable

El objetivo de utilizar la técnica de acoplamiento para acotar las distancias de variación y separación total, es obtener un acoplamiento que genere un tiempo de acoplamiento T que se asocie con algún tiempo aleatorio (de paro) del cual se conoce el comportamiento de su distribución ó entonces producir un tiempo T donde se pueda analizar la distribución a manera de producir un resultado interpretable.

A lo largo de la tesis se muestra la búsqueda y el encuentro de un tiempo de corte para cadenas de Markov utilizando la técnica de acoplamiento. Haciéndolo de la siguiente manera: primero conociendo bien el problema y presentando un tiempo de acoplamiento cuya distribución es una suma de distribuciones geométricas con parámetros conocidos. Después se usa la relación existente entre este tiempo de acoplamiento y el tiempo necesario para coleccionar todos los cupones en el problema del coleccionador para conseguir el comportamiento asintótico de la distribución del tiempo de acoplamiento. En nuestro caso T es aportada por el problema del coleccionador de cupones.

Para otros problemas se puede utilizar la relación con otros tiempos aleatorios, por ejemplo, para un proceso de nacimiento y muerte con ciertas características se puede obtener un tiempo de absorción; del problema de las esposas, un tiempo de paro óptimo; de un problema de filas, un tiempo de espera, entre otros.

4.5 Otros barajeos

En el capítulo 3 se describen otros barajeos. Ahora se agrega que gracias a la aplicación de un acoplamiento se pueden establecer su tiempo aproximado de corte. (Ver Aldous (1983))

Así se tiene que para:

Top to random, bottom to random. $T \cong N \log N$.

Transposing. $C_1 \leq T \leq C_2 N^3 \log N$, para algunas C_1, C_2 constantes.

Random transpositions. $\frac{1}{2}N \log N \leq T \leq CN^2$, para alguna C constante.

4.6 Monte Carlo vía cadenas de Markov

Las técnicas de Monte Carlo han tenido un gran éxito a últimas fechas. Ellas auxilian en el análisis de modelos estadísticos complejos buscando con ello simular datos o conjuntos de datos faltantes o desconocidos.

El problema básico es simular datos a partir de una distribución de enteros que presenta una forma del estilo de la del siguiente ejemplo:

$$\pi(x) + \frac{1}{c}h(x),$$

donde $h(\cdot)$ tiene una forma producto y c es no calculable desde el punto de vista computacional. De este modo, lo que los métodos de Monte Carlo vía cadenas de Markov hace es construir una cadena de Markov X , ergódica, cuya distribución estacionaria es $\pi(\cdot)$. De este modo, si se hace un número suficiente de iteraciones de X , los valores simulados provienen de la distribución $\pi(\cdot)$. Es importante mencionar las dos variantes de mayor uso:

- Algoritmo Metropolis-Hastings
- Muestreo de Gibbs

Es importante señalar que el mayor problema al que se enfrentan estas técnicas es saber el tiempo de CONVERGENCIA al estado estacionario de la cadena de Markov utilizada. Este tiempo de convergencia es conocido entre los usuarios del método de Monte Carlo como el período de “calentamiento”. Explicándolo de una manera un poco burda, se tiene que después de crear el modelo en cadena de Markov a usar, se simulará iterando cuantas veces sea necesario hasta que la cadena alcance el estado de equilibrio, pero... ¿cuántas iteraciones son las necesarias?, ¿dónde parar? En términos de lo estudiado durante esta tesis ¿cuándo la cadena de Markov ha convergido al estado estacionario?, ésta es precisamente la relación que tiene el **acoplamiento** con el **método Monte Carlo**. El utilizar el acoplamiento resuelve esta pregunta, ahorrando tiempo de observación en el laboratorio de cómputo para descubrir en que n^* exacta o aproximada (cota) se iniciará la fase de estabilización. Es de suponerse que dependiendo del problema se diseñará el acoplamiento a usar y que no siempre se tendrá un análisis sencillo, no dejando de ser con esto una técnica muy útil.

Capítulo 5

Apéndice

“Preferiría no decir nada antes de que expresarme débilmente.”

Millet

En esta sección, se va a dar algunos resultados elementales que fueron utilizados durante el desarrollo de esta tesis.

5.1 Desigualdad de Chebyshev

La desigualdad de Chebyshev es una desigualdad de segundo orden derivada de la desigualdad de Markov. Primero se demostrará la desigualdad de Markov y después la utilización de la desigualdad de Markov para demostrar la desigualdad de Chebyshev.

Teorema 5.1.1. *Sea X una variable aleatoria asumiendo valores positivos, entonces para todo $a > 0$ se tiene*

$$P(X > a) \leq \frac{E(X)}{a}$$

Demostración. Sea X variable aleatoria continua. Para una variable aleatoria discreta, la demostración es similar utilizando sumas en lugar de integrales.

Sean f y P la función de densidad y la función de probabilidad de X respectivamente, entonces

$$\begin{aligned}
 E(X) &\stackrel{(1)}{=} \int_0^{\infty} xf(x)dx \\
 &= \int_0^a xf(x)dx + \int_a^{\infty} xf(x)dx \\
 &\stackrel{(2)}{\geq} \int_a^{\infty} xf(x)dx \\
 &\stackrel{(3)}{\geq} \int_a^{\infty} af(x)dx = a \int_a^{\infty} f(x)dx \\
 &\stackrel{(4)}{=} aP(X \geq a)
 \end{aligned}$$

y por tanto

$$P(X > a) \leq \frac{E(X)}{a}$$

(1). Definición de esperanza.

(2). Dado que $x \geq 0$, entonces $\int_0^a xf(x)dx \geq 0$.

(3). Porque $x \geq a$.

(4). Porque $f(x)$ es la densidad de X . □

Teorema 5.1.2. Desigualdad de Chebyshev

Sea X una variable aleatoria con segundo momento finito, es decir, $E[X^2] < \infty$ y varianza $\text{Var}(X)$, entonces para todo $k > 0$

$$P(|X - E(X)| > k) \leq \frac{\text{Var}(X)}{k^2}$$

Demostración. Defina $Y = (X - E[X])^2$ y sea $f_Y(\cdot)$ la función de densidad de la variable aleatoria Y . Note que Y es una variable aleatoria positiva con esperanza finita, entonces aplicando la desigualdad de Markov a Y con $a = k^2$ se tiene

$$P(Y > k^2) \leq \frac{E(Y)}{k^2}$$

Ahora se tiene que por definición $Y = [X - E(X)]^2$ y por tanto,

$$E(Y) = E([X - E(X)]^2) = \text{Var}(X),$$

donde sigue que

$$P(|x - E(X)| > k) \leq \frac{\text{Var}(X)}{k^2}$$

como se quería demostrar. \square

5.2 Distribución geométrica

Definición 5.2.1. Se dice que una variable aleatoria X tiene **distribución geométrica** con probabilidad de éxito $p \in (0, 1)$ si

$$P(X = x) = \begin{cases} p(1-p)^{x-1} & , \text{ si } x = 1, 2, \dots \\ 0 & \text{ en otro caso} \end{cases}$$

Lema 5.2.2. Para X una variable aleatoria geométrica con parámetro p se tiene que $E(X) = \frac{1}{p}$ y $\text{Var}(X) = \frac{1-p}{p^2}$

Demostración.

$$\begin{aligned} E[X] &= \sum_{x=1}^{\infty} x(1-p)^{x-1}p \\ &= p \sum_{x=1}^{\infty} x(1-p)^{x-1} \\ &\stackrel{(1)}{=} p \left(\sum_{x=1}^{\infty} \frac{-d}{dp} (1-p)^x \right) \\ &\stackrel{(2)}{=} p \frac{d}{dp} \left(- \sum_{x=1}^{\infty} (1-p)^x \right) \\ &= -p \frac{d}{dp} \left[\frac{1}{1 - (1-p)} - 1 \right] \\ &= -p \frac{d}{dp} \left(\frac{1-p}{p} \right) \\ &= \frac{1}{p} \end{aligned}$$

- (1). $x(1-p)^{x-1} = \frac{-d}{dp}(1-p)^x$.
 (2). Dado que la serie es convergente.

Ahora, se pasa al calculo de la varianza de X .

$$\text{Var}(X) = E(X^2) - E^2(X).$$

Ahora, se tiene que

$$E^2(X) = \sum_{x=1}^{\infty} x^2 p(1-p)^{x-1} = p \sum_{x=1}^{\infty} x^2 (1-p)^{x-1}$$

Si se renombra $q = 1 - p$, entonces se tiene que

$$E(X^2) = p \sum_{x=1}^{\infty} x^2 q^{x-1}.$$

Considere las siguientes series,

$$S_1(x) = \sum_{x=1}^{\infty} x^2 q^{x-1}$$

y

$$S_2(x) = q \sum_{x=1}^{\infty} x^2 q^{x-1}.$$

Note que

$$\begin{aligned}
 S_1(x) - S_2(x) &= \sum_{x=1}^{\infty} x^2 q^{x-1} (1-q) \\
 &= p \sum_{x=1}^{\infty} x^2 q^{x-1} \\
 E(X^2) &= \sum_{x=1}^{\infty} x^2 q^{x-1} (1-q) \\
 &= \sum_{x=1}^{\infty} x^2 q^{x-1} - \sum_{x=1}^{\infty} x^2 q^x \\
 &\stackrel{(1)}{=} \sum_{y=0}^{\infty} (y+1)^2 q^y - \sum_{x=1}^{\infty} x^2 q^x \\
 &= 1 + \sum_{x=1}^{\infty} (x+1)^2 q^x - \sum_{x=1}^{\infty} x^2 q^x \\
 &= 1 + \sum_{x=1}^{\infty} q^x [(x-1)^2 - x^2] \\
 &= 1 + \sum_{x=1}^{\infty} q^x [2x+1] \\
 &= \sum_{x=1}^{\infty} q^x 2x + \left(1 + \sum_{x=1}^{\infty} q^x\right) \\
 &= \sum_{x=1}^{\infty} q^x 2x + \left(\sum_{x=0}^{\infty} q^x\right)
 \end{aligned}$$

$$\begin{aligned}
 &\stackrel{(2)}{=} \frac{2q}{(1-q)^2} + \frac{1}{1-q} \\
 &= \frac{1}{p^2} (2q+p) = \frac{1}{p^2} [2(1-p)+p] \\
 &= \frac{1}{p^2} [2-p]
 \end{aligned}$$

Bibliografía

- [1] Ash, R.B. (1972) *Real analysis and probability* Academic Press.
- [2] Aldous, D. (1983) *Random walks on finite groups and rapidly mixing Markov Chains*, Seminaire de Probabilités XVII, Springer Lecture Notes in Mathematics, Vol. 986, Springer-Verlag, Berlin.
- [3] Aldous, D. y Diaconis, P. (1986) *Shuffling cards and stopping times*, Amer. Mathematics Monthly 93, 333-348.
- [4] Aldous, D. y Diaconis, P. (1987) *Strong uniform times and finite random walks*, Advances in applied mathematics 8, 69-97.
- [5] Aldous, D. (1989) *Probability approximations via the Poisson Clumping*, Applied mathematical sciences 77, Springer Verlag.
- [6] Boneh, A. y Hofri, M. *The coupon collector problem revisited - a survey of engineering problems and computational methods*, Comm. Stat. Stochastic Models 13, 39-66.
- [7] Bosq, D. y Nguyes, Hung T. (1996) *A course in Stochastic Processes, stochastic models and statistical inferences*, Kluwer Academic Publishers.
- [8] Csörgö, S. *A rate of convergence for the coupon collectors*, preimpreso
- [9] Feller, W. (1968) *An introduction to probability theory and its applications*, 3 edn, vol. I, Wiley and Sons.
- [10] Fill, J.A. (1996) *An exact formula for the move-to-front rule for self-organizing lists*, Reprinted from Journal Theoretical Probability, Vol. 9, No. 1

- [11] Goldstein, S. (1979) *Maximal coupling*, Z. Wahrsch. Verw. Gebiete 46, 193-204.
- [12] Griffeath, D. (1975) *A maximal coupling for Markov chain*, Z. Wahrsch. Verw. Gebiete 31, 95-106
- [13] Isaacson, Dean L. y Madsen, Richard W. (1976) *Markov Chains theory an applications*, John Wiley and Sons, Inc.
- [14] Karlin, Samuel (1968) *A first course in Stochastic Processes*, primera edición, Academic Press, New York.
- [15] Lindvall, Torgny (1992) *Lectures on the Coupling Method*, Wiley-Interscience Publication, Jonh Wiley and Sons, Inc.
- [16] Parzen, Emanuel (1972) *Procesos estocásticos*, Madrid, Paraninfo.
- [17] Ross, Sheldon M. (1972) *Introduction to Probability Models*, Academic Press, Inc.
- [18] Ross, Sheldon M. (1996) *Stochastic Processes*, 2 edn. Wiley Series in Probability and Mathematical Statistics, 1996
- [19] Rodrigues, Eliane R. (1995) *Convergence to stationary state for a Markov move to front scheme*, J. Appl. Prob. 32, 768-776.
- [20] Rodrigues, Eliane R. (1995) *The performance of the move to front scheme under some particular forms of Markov request*, J. Appl. Prob. 32, 1089-1102.
- [21] Takács, L. (1978) *Stochastic Processes. Problems and solutions*, Chapman and Hall.
- [22] Taylor, Howard M. y Karlin, Samuel (1984) *Introduction to stochastic modeling*, Academic Press, Inc.
- [23] Thorisson, H. (1986) *On maximal and distributional coupling*, Ann. Probab, 14, 873-876