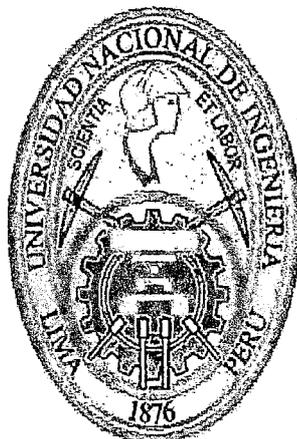


UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE CIENCIAS
Escuela Profesional de Matemática



Tesis para optar el Título Profesional de
Licenciado en Matemática

**ESTIMACIÓN PUNTUAL Y POR
INTERVALOS PARA CADENAS DE
MARKOV**

Presentada por:

JOSÉ LUIS ESPINOZA MELGAREJO

Asesor:

MG. WILLIAM ECHEGARAY CASTILLO

LIMA-PERÚ

2014

Digitalizado por:

**Consortio Digital del
Conocimiento MebLatam,
Hemisferio y Dalse**

Este trabajo está dedicado, en primer lugar, a tres personas que son importantes en mi vida, mis sobrinos Mateo, Luana y Cielo. También a mis padres Marcos y Beatriz, por haberme brindado una educación que yo considero de calidad, y a mis hermanos mayores Patricia y Roberto, por sus consejos y ayuda cuando más los necesité.

También quería dar un agradecimiento especial a mi asesor, el profesor William Echeagaray Castillo, director de la Escuela Profesional de Matemática por su valioso apoyo y paciencia para la elaboración de esta tesis y, además, por haber contribuido a mi formación académica durante mi época de estudiante en esta facultad.

Resumen

En el presente trabajo presentamos un estudio sobre un proceso estocástico importante: las cadenas de Markov. Realizaremos el proceso de inferencia estadística (estimación puntual y por intervalos) sobre las probabilidades de transición de dicha cadena, tanto para el caso paramétrico como para el no paramétrico. Una vez hecha la estimación por intervalos verificaremos, mediante un proceso de simulación, las probabilidades de cobertura asociadas a dichos intervalos de verosimilitud (nivel de confianza real) y la compararemos con lo teóricamente indicado (nivel de confianza nominal). Para el caso no paramétrico, nos valdremos de un ejemplo de aplicación real consistente en datos de precipitación diaria que fueron tomados de una estación meteorológica del Senamhi, ubicada en la provincia de Pasco, en el distrito de Chaupimarca.

Índice general

Introducción	8
1. Conceptos Preliminares	11
1.1. Variables aleatorias	11
1.2. Teoremas de convergencia y ley de los grandes números	16
1.3. Nociones de teoría de la medida y cálculo	17
1.4. La distribución multinomial	19
1.5. Multiplicadores de Lagrange	20
2. Teoría de Cadenas de Markov	22
2.1. Conceptos básicos	22
2.2. Caminata aleatoria simple	23
2.2.1. La distribución de la fortuna	24
2.2.2. El problema de la ruina del jugador	25
2.3. Cadenas de Markov	27
2.4. Cálculos con cadenas de Markov	33
2.5. Clasificación de estados de una cadena de Markov	40
2.5.1. Comunicación	40
2.5.2. Periodo	42
2.5.3. Primeras visitas	44
2.5.4. Recurrencia y transitoriedad	46
2.5.5. Tiempo medio de recurrencia	52
2.5.6. Número de visitas	53
2.5.7. Distribución límite de una cadena de Markov	58
3. Inferencia Estadística	78
3.1. Función de verosimilitud	78
3.2. Función de verosimilitud relativa	79
3.3. Propiedades de la verosimilitud	80

3.3.1. No aditividad de las verosimilitudes	80
3.3.2. Combinación de experimentos	80
3.3.3. Invarianza funcional	80
3.4. Aproximación continua de la función de verosimilitud y la verosimilitud exacta	81
3.5. Intervalos de verosimilitud	82
3.6. Probabilidad de cobertura de intervalos de verosimilitud	83
4. Verosimilitud para Cadenas de Markov	93
4.1. Caso paramétrico	93
4.1.1. Simulación de una cadena de Markov paramétrica	96
4.2. Caso no paramétrico	99
4.3. Ejemplo de aplicación real	101
4.3.1. Determinación del orden de una cadena de Markov de dos estados . . .	103
4.3.2. Simulación de una cadena de Markov no paramétrica	105
Conclusiones	110
Apéndice	111
Bibliografía	145

Índice de tablas

3.1. Confianza aproximada de $IV(c)$ cuando θ es unidimensional	88
3.2. Resumen de inferencias para θ	91
4.1. Resumen de inferencias para el caso paramétrico	95
4.2. Resumen de inferencias para la simulación paramétrica	98
4.3. Probabilidades de cobertura a los diferentes niveles de plausibilidad	98
4.4. Valores del AIC y BIC para diferentes órdenes	105
4.5. Intervalos de verosimilitud para \widehat{p}_{11}	106
4.6. Intervalos de verosimilitud para \widehat{p}_{12}	106
4.7. Intervalos de verosimilitud para \widehat{p}_{21}	106
4.8. Intervalos de verosimilitud para \widehat{p}_{22}	106
4.9. Intervalos de verosimilitud para \widehat{p}_{11} de la cadena simulada	107
4.10. Intervalos de verosimilitud para \widehat{p}_{12} de la cadena simulada	108
4.11. Intervalos de verosimilitud para \widehat{p}_{21} de la cadena simulada	108
4.12. Intervalos de verosimilitud para \widehat{p}_{22} de la cadena simulada	108
4.13. Probabilidades de cobertura a los diferentes niveles de plausibilidad	109

Índice de figuras

2.1. Clases de comunicación para el ejemplo 2.5.1	41
3.1. Intervalos de verosimilitud	91
4.1. Pluviómetro tipo Hellman modelo t027o TP1K-HLM	102

Introducción

Uno de los principales recursos naturales del que disponemos en nuestra sociedad es el agua, imprescindible para cualquier actividad humana, pero en especial para las actividades agrícolas, sobre todo en la serranía de nuestro país. Por ello es importante conocer la probabilidad diaria de ocurrencia o de no precipitación, así como la duración de las rachas lluviosas y secas, pues constituyen una valiosa aportación para poder gestionar los recursos hídricos de manera adecuada. En climatología, se entiende por racha lluviosa a si se registra, o no, alguna precipitación durante una serie de días consecutivos. Por otro lado, si se tiene en cuenta, además, la cantidad de lluvia precipitada, se estará considerando una secuencia como lluviosa no solo atendiendo a la presencia del líquido elemento sino, también al hecho de que el volumen de la precipitación registrada cada día sea superior a un valor o umbral de corte determinado (dicho umbral se mide en mm/día).

Los procesos estocásticos estudian y caracterizan fenómenos aleatorios como el descrito anteriormente, que evoluciona con el tiempo. Uno de estos procesos es el proceso de Markov que plantea que los valores de los procesos futuros dado los valores de los procesos actuales no dependen de los valores pasados. Una cadena de Markov a tiempo discreto es un proceso de Markov. Ejemplos de ellos son la caminata aleatoria simple, teoría de colas, ruina del jugador, etc.

Los procesos estocásticos sirven para conocer el desarrollo futuro de procesos que se realizan a lo largo del tiempo, para predecir el comportamiento de fenómenos físicos, económicos, atmosféricos, etc. Las cadenas de Markov constituyen un caso particular de estos tipos de procesos. Estas fueron introducidas por el matemático ruso Andréi Markov alrededor del año 1905. Su intención era crear un modelo probabilístico para analizar la frecuencia con la que aparecen las vocales en poemas y textos literarios. El éxito del modelo propuesto radica en que es lo suficientemente complejo como para describir ciertas características no triviales de algunos sistemas, pero al mismo tiempo es lo suficientemente sencillo para ser analizado matemáticamente. Las cadenas de Markov pueden aplicarse a una amplia gama de fenómenos científicos y sociales, y se cuenta con una teoría matemática extensa al respecto. En nuestro país no se encuentran trabajos en donde se utilice el principio de máxima verosimilitud para

estimar las probabilidades de transición de una cadena de Markov, debido tal vez a la poca difusión de este método en trabajos de investigación (el cual resulta ser una buena alternativa para estimación de parámetros de este tipo). La estimación por máxima verosimilitud es un método habitual para ajustar un modelo y encontrar sus parámetros. Para ello nos valemos de una función conocida como *función de verosimilitud*, la cual es proporcional a la función de probabilidad conjunta de las variables aleatorias que conforman la muestra. Por ejemplo, supongamos que se colocan en una urna, n bolas numeradas de 1 a n y que una de ellas se extrae al azar. Si se desconoce n , su estimador de máxima verosimilitud (el valor que maximiza la función de verosimilitud) es el número m que aparece en la bola extraída: la función de verosimilitud es 0 para $n < m$ y $1/n$ para $n \geq m$, que alcanza su máximo cuando $n = m$. Sin embargo, fuera de nuestras fronteras sí se encuentran trabajos al respecto, como por ejemplo, Anderson y Goodman ([1]) proponían que el estimador de máxima verosimilitud y sus distribuciones asintóticas son obtenidos para las probabilidades de transición en una cadena de Markov de orden arbitrario cuando existen observaciones repetidas de la cadena. Las pruebas de radio verosimilitud y las pruebas chi-cuadrado de la forma usada en las tablas de contingencia son obtenidas probando las siguientes hipótesis: (a) que las probabilidades de transición de una cadena de primer orden son constantes, (b) que en caso que las probabilidades de transición son constantes, ellas son números especificados, y (c) que el proceso es una cadena de Markov de u -ésimo orden contra la alternativa que esta es de orden r -ésimo pero no u -ésimo. En caso que $u = 0$ y $r = 1$, el caso (c) resulta en la prueba de la hipótesis nula de que las observaciones en los puntos de tiempo sucesivos son estadísticamente independientes contra la hipótesis alternativa que las observaciones son de la forma de una cadena de Markov de primer orden. Craig y Sendi ([6]) utilizan un modelo de Markov de tiempo discreto para investigar programas de tratamiento y protocolos de atención de salud para enfermedades crónicas. La matriz de transición, la cual describe la progresión natural de la enfermedad, es a menudo estimada de una observación de cohorte en intervalos en común. Aquí se resumen métodos para obtener el estimador de máxima verosimilitud de la matriz de transición. En la tesis de Skuriat-Olechnowska ([32]) se considera un modelo de deterioro de puentes en Holanda. Este modelo está basado en una cadena de Markov con una condición de escala discreta. Las transiciones en la cadena de Markov ocurren en unidades de tiempo discreto, usualmente todo el año. La condición de deterioro de los puentes es representada por una número de 0 a 6, donde 6 es perfecto y 0 extremadamente malo, donde cada número representa un estado de la cadena. Existen otros trabajos que tratan este tema como por ejemplo en el trabajo de Salazar *et al.* ([30]) se considera un modelo de tres estados donde uno de ellos es absorbente y las observaciones registradas para un mismo sujeto siguen un proceso de Markov. El método de estimación de parámetros se basa en la función de verosimilitud que se optimiza mediante las ecuaciones de Kolmogorov hacia adelante junto con el algoritmo de Newton-Raphson. En

otro estudio publicado por Tovar y Salazar ([34]) se considera un modelo de $k + 2$ estados donde dos de ellos son absorbentes y k son transitorios obteniendo la función de verosimilitud y a partir de ella se hace la maximización de dicha función mediante los métodos numéricos de la cuadratura de Gauss y el algoritmo de Newton-Raphson.

El presente trabajo se encuentra dividido en cuatro capítulos. En el primer capítulo se presentan brevemente algunos conceptos preliminares como variables aleatorias, teoremas de convergencia y otras nociones previas de teoría de la medida, cálculo y optimización.

En el segundo capítulo se dan definiciones importantes acerca de las cadenas de Markov, las cuales serán importantes para los capítulos siguientes. También se mostrarán teoremas, proposiciones, corolarios y ejemplos de cadenas de Markov.

En el tercer capítulo se mostrará cómo hacer inferencia estadística por el método de máxima verosimilitud y su aplicación mediante un ejemplo.

En el cuarto capítulo se desarrollará la metodología propuesta de hacer inferencia estadística sobre las probabilidades de transición de una cadena de Markov considerando dos casos: el caso paramétrico y el no paramétrico. También, mediante una simulación, se verificará la propiedad de verosimilitud de cobertura de intervalos. Finalmente se mostrará un ejemplo de aplicación real de una cadena de días lluviosos o secos.

Los objetivos de la presente tesis son los siguientes:

- (1) Aplicar el método de verosimilitud sobre las probabilidades de transición de una cadena de Markov.
- (2) Obtener simulaciones de observaciones de una cadena de Markov y verificar la propiedad de verosimilitud de cobertura de intervalos.
- (3) Predecir la ocurrencia de un día lluvioso o seco.
- (5) Probar que existe una relación entre la estadística y la probabilidad, ante una misma situación problemática.

Capítulo 1

Conceptos Preliminares

En este capítulo estableceremos algunos conceptos y resultados sobre variables aleatorias, funciones de distribución, esperanza matemática, teoremas de convergencia, así como algunos resultados de teoría de la medida y cálculo, que usaremos en los capítulos posteriores.

1.1. Variables aleatorias

En esta parte introduciremos algunas nociones básicas de la teoría de la probabilidad necesarias para la definición de procesos estocásticos, y por ende, de las cadenas de Markov.

En la práctica ocurren dos tipos de experimentos o fenómenos: los determinísticos, los cuales son determinadas acciones en las que se puede predecir qué es lo que va a ocurrir con exactitud, y los aleatorios, los cuales son acciones en las que no sabemos qué ocurrirá al ser realizados, ya que pueden obtenerse cualquiera de varias alternativas, aun bajo las mismas condiciones en que se realizan estas acciones. A continuación daremos algunas definiciones que nos permitirán adentrarnos en el desarrollo de esta teoría.

Definición 1.1.1 *Dado un fenómeno aleatorio, al conjunto de puntos de los posibles resultados se le llama **espacio muestral** y los denotamos por Ω .*

Definición 1.1.2 *Sea un fenómeno aleatorio con espacio muestral Ω , una **variable aleatoria** es una función X cuyo dominio es Ω y contradominio es \mathbb{R} , así $X : \Omega \rightarrow \mathbb{R}$.*

Adicionalmente a estas definiciones, llamaremos evento a cualquier subconjunto del espacio muestral y definiremos a la clase \mathcal{F} como la colección de todos los eventos de Ω .

Con las definiciones anteriores, la medida de probabilidad se puede ver como sigue. Sea \mathcal{F} la clase de eventos de Ω , entonces para cada evento $A \in \mathcal{F}$ se le asigna un número $P(A) \in [0, 1]$.

Este número es la fracción esperada de ocurrencia del evento A , a lo largo de una serie de experimentos donde A o A^c son observados.

Algunas propiedades elementales de la medidas de probabilidad son resumidas a continuación.

i) Se cumple que

$$P(\Omega) = 1 \text{ y } P(\emptyset) = 0$$

ii) Para eventos $A, B \in \mathcal{F}$

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

iii) Si A y B son disjuntos, es decir, $A \cap B = \emptyset$, entonces

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

Además, de la propiedad iii) se sigue que

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

La relación entre las variables aleatorias y probabilidad puede ser caracterizada por ciertas cantidades numéricas. Ahora consideremos algunas de estas.

Definición 1.1.3 *La colección de las probabilidades*

$$F_X(x) = P(X \leq x) = P(\{\omega : X(\omega) \leq x\}). \quad x \in \mathbb{R}, \omega \in \Omega$$

es la *función de distribución de probabilidad acumulada* F_X de X .

Utilizando la definición anterior 1.1.3 podemos determinar probabilidades en términos de la función de distribución de probabilidad acumulada, como

1. $P(\{\omega : a < X(\omega) \leq b\}) = F_X(b) - F_X(a)$ con $a < b$
2. $P(X = x) = F_X(x) - \lim_{h \rightarrow 0} F_X(x - h)$

Con estas probabilidades se puede aproximar la probabilidad del evento $\{\omega : X(\omega) \in B\}$ para cualquier conjunto complicado $B \subset \mathbb{R}$. Daremos otra definición que nos servirá también para calcular la probabilidad de cualquier evento, pero primero daremos algunos previos.

Definición 1.1.4 Una σ -álgebra \mathcal{F} sobre Ω es una familia de subconjuntos de Ω que satisfacen las siguientes condiciones:

- i. $\emptyset, \Omega \in \mathcal{F}$
- ii. Si $A \in \mathcal{F}$, entonces $A^c \in \mathcal{F}$
- iii. Si $A_1, A_2, \dots \in \mathcal{F}$, entonces $\bigcup_{n=1}^{\infty} A_i \in \mathcal{F}$

Ahora supongamos que \mathcal{C} es una familia de subconjuntos de Ω , pero no necesariamente una σ -álgebra. Añadiendo más conjuntos a \mathcal{C} , se puede obtener siempre una σ -álgebra, por ejemplo el conjunto potencia $P(\Omega)$. Por razones matemáticas, $P(\Omega)$, en general, resulta demasiado grande. Además se puede probar que: "Para una familia dada \mathcal{C} de subconjuntos de Ω , existe una σ -álgebra más pequeña $\sigma(\mathcal{C})$ en Ω que contiene a \mathcal{C} ". Denotaremos a $\sigma(\mathcal{C})$ como la σ -álgebra generada por \mathcal{C} .

Tomemos $\Omega = \mathbb{R}$ y $\mathcal{C}^{(1)} = \{ \langle a, b \rangle : -\infty < a < b < \infty \}$. La σ -álgebra $\mathcal{B}^1 = \sigma(\mathcal{C}^{(1)})$ contiene conjuntos muy generales de \mathbb{R} . Esta se denomina σ -álgebra de Borel y sus elementos son los conjuntos de Borel. También se puede introducir la σ -álgebra de los conjuntos de Borel n -dimensionales $\mathcal{B}^n = \sigma(\mathcal{C}^{(n)})$, donde $\Omega = \mathbb{R}^n$ y $\mathcal{C}^{(n)} = \{ \langle \mathbf{a}, \mathbf{b} \rangle : -\infty < a_i < b_i < \infty, \mathbf{a} = (a_1, a_2, \dots, a_n), \mathbf{b} = (b_1, b_2, \dots, b_n) \}$. Los conjuntos $\langle \mathbf{a}, \mathbf{b} \rangle$ se denominan rectángulos. Todo conjunto "razonable" de \mathbb{R}^n es un conjunto de Borel, por ejemplo, bolas, esferas, curvas suaves, superficies, etc. y también los conjuntos abiertos y cerrados. En orden a demostrar que un subconjunto $B \subset \mathbb{R}^n$ es un conjunto de Borel, es necesario obtener B mediante un número numerable de operaciones \cap, \cup o c , que operen sobre los rectángulos.

Por ejemplo, se puede demostrar para $n = 1$ que todo subconjunto de puntos $\{a\}$, con $a \in \mathbb{R}$, es un conjunto de Borel. También se puede comprobar que los intervalos $\langle a, b \rangle$, $[a, b]$, $\langle -\infty, a \rangle$, $\langle b, \infty \rangle$ son conjuntos de Borel.

Ahora sí, estamos en condiciones de dar la siguiente definición para calcular la probabilidad de cualquier evento.

Definición 1.1.5 La colección de las probabilidades

$$P_X(B) = P(X \in B) = P(\{\omega : X(\omega) \in B\})$$

es la *distribución* de X para $B \subset \mathbb{R}$.

La distribución P_X y la función de distribución de probabilidad acumulada F_X son nociones equivalentes en el sentido de que ambas pueden ser usadas para calcular la probabilidad de cualquier evento $\{X \in B\}$.

Una función de distribución de probabilidad es continua o discreta (tiene saltos). Consideremos primero el caso especial cuando la función de distribución F_X es una función discreta.

Definición 1.1.6 Sea X una variable aleatoria discreta con función de probabilidad p_k , se define su función de distribución de probabilidad acumulada por

$$F_X(x) = \sum_{k: x_k \leq x} p_k, \quad x \in \mathbb{R} \quad (1.1)$$

donde

$$0 \leq p_k \leq 1 \text{ para todo } k \text{ y } \sum_{k=1}^{\infty} p_k = 1$$

y

$$p_k = P(X = x_k)$$

La función de distribución de probabilidad acumulada (1.1) y la correspondiente distribución se dice que son discretas. Una variable aleatoria con función de distribución de probabilidad acumulada (1.1) es una **variable aleatoria discreta**.

En contraste a las distribuciones y variables aleatorias discretas, la función de distribución de una variable aleatoria continua no tiene saltos, por lo tanto, $P(x) = 0$ para todo x , o equivalentemente

$$\lim_{h \rightarrow 0} F_X(x+h) = F_X(x), \quad \forall x \quad (1.2)$$

es decir, tal variable aleatoria asume cualquier valor con probabilidad 0. Una variable aleatoria continua obtiene su nombre de la propiedad de continuidad (1.2) de la función de distribución de probabilidad acumulada F_X .

Definición 1.1.7 Sea X una variable aleatoria continua con función de densidad de probabilidad f_X , se define su función de distribución de probabilidad acumulada por

$$F_X(x) = \int_{-\infty}^x f_X(y) dy, \quad x \in \mathbb{R}$$

donde, la función de densidad de probabilidad cumple con

$$f_X(x) \geq 0, \forall x \in \mathbb{R} \text{ y } \int_{-\infty}^{\infty} f_X(x) dx = 1$$

Por otro lado, algunas características interesantes de una variable aleatoria X son la esperanza $E(X)$, la varianza $Var(X)$ y los momentos $E(X^l)$.

A continuación definiremos estas características tanto para variables aleatorias discretas como continuas.

Definición 1.1.8 Sea X una variable aleatoria continua con función de densidad de probabilidad f_X .

i) La esperanza o valor medio de X está dada por

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

ii) La varianza de X está definida como

$$\sigma_X^2 = Var(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

iii) El l -ésimo momento de X para $l \in \mathbb{N}$ está definido como

$$E(X^l) = \int_{-\infty}^{\infty} x^l f_X(x) dx$$

iv) Para una función real g y la esperanza de $g(X)$ está dada por

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

Similarmente se define

Definición 1.1.9 Sea X una variable aleatoria discreta con probabilidades $p_k = P(X = x_k)$.

i) La esperanza o valor medio de X está dada por

$$\mu_X = E(X) = \sum_{k=1}^{\infty} x_k p_k$$

ii) La varianza de X está definida como

$$\sigma_X^2 = Var(X) = \sum_{k=1}^{\infty} (x_k - \mu_X)^2 p_k$$

iii) El l -ésimo momento de X para $l \in \mathbb{N}$ está definido como

$$E(X^l) = \sum_{k=1}^{\infty} x_k^l p_k$$

iv) Para una función real g y la esperanza de $g(X)$ está dada por

$$E[g(X)] = \sum_{k=1}^{\infty} g(x_k) p_k$$

Podemos considerar la esperanza μ_X como el *centro de gravedad* de la variable aleatoria X , es decir, los valores aleatorios $X(\omega)$, con $\omega \in \Omega$, que están concentrados alrededor del valor no aleatorio μ_X . La extensión o dispersión de los valores aleatorios $X(\omega)$ alrededor de la esperanza μ_X es descrita por la varianza σ_X^2 y la desviación σ_X .

1.2. Teoremas de convergencia y ley de los grandes números

Teorema 1.2.1 (De convergencia dominada) Si X, X_1, X_2, \dots son variables aleatorias, donde $\{X_n\} \rightarrow X$ con probabilidad 1, y si existe una variable aleatoria Y con $|X_n| \leq Y$ para todo n con $E(Y) < \infty$, entonces

$$\lim_{n \rightarrow \infty} E(X_n) = E(X)$$

Demostración (ver [28, pág. 104])

Dado un espacio de probabilidad $(\Omega, \mathcal{F}, \mathcal{P})$ y los eventos $A, A_1, A_2, \dots \in \mathcal{F}$, la notación $\{A_n\} \nearrow A$ significa que $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$, y $\bigcup_n A_n = A$. Es decir, los eventos A_n aumentan hasta A . Similarmente, la notación $\{A_n\} \searrow A$ significa que $\{A_n^c\} \nearrow A^c$, o equivalentemente que $A_1 \supseteq A_2 \supseteq A_3 \dots$, y $\bigcap_n A_n = A$. Es decir, los eventos A_n decrecen hasta A .

Teorema 1.2.2 (De convergencia monótona) Supongamos que X_1, X_2, \dots son variables aleatorias con $E(X_1) > -\infty$, y $\{X_n\} \nearrow X$. Entonces X es una variable aleatoria y

$$\lim_{n \rightarrow \infty} E(X_n) = E(X)$$

Demostración (ver [28, pág. 46])

Teorema 1.2.3 (Ley débil de los grandes números) Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, donde cada una de ellas tiene $E(X_i) = \mu$ y $\text{Var}(X_i) \leq \sigma^2$. Entonces para todo $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n}(X_1 + X_2 + \dots + X_n) - \mu \right| \geq \epsilon \right) = 0$$

Es decir, el promedio $\frac{1}{n}(X_1 + X_2 + \dots + X_n)$ converge en probabilidad a μ .

Demostración ([28, pág. 60])

Teorema 1.2.4 (Ley fuerte de los grandes números (primera versión)) Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes, donde cada una de ellas tiene $E(X_i) = \mu$ y $E((X_i - \mu)^4) \leq a < \infty$. Entonces para todo $\epsilon > 0$

$$P \left(\lim_{n \rightarrow \infty} \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \mu \right) = 1$$

Es decir, el promedio $\frac{1}{n}(X_1 + X_2 + \dots + X_n)$ converge casi con toda seguridad a μ .

Demostración (ver [28, págs. 60-61])

Teorema 1.2.5 (Ley fuerte de los grandes números (segunda versión)) Sea X_1, X_2, \dots una sucesión de variables aleatorias independientes e idénticamente distribuidas, donde cada una de ellas tiene $E(X_i) = \mu$. Entonces para todo $\epsilon > 0$

$$P \left(\lim_{n \rightarrow \infty} \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \mu \right) = 1$$

Es decir, el promedio $\frac{1}{n}(X_1 + X_2 + \dots + X_n)$ converge casi con toda seguridad a μ .

Demostración (ver [28, págs. 62-64])

1.3. Nociones de teoría de la medida y cálculo

El conjunto potencia de un conjunto X dado, definido como la colección de todos los subconjuntos de X , es una σ -álgebra en X . Se suele denotar al conjunto potencia de X por 2^X .

Si $E \subset X$, denotamos por $\mu_C(E)$ al número de elementos de E (pudiendo ser finito o infinito). Esta medida μ_C también es conocida como la *medida de contar*.

Teorema 1.3.1 (Teorema de Fubini) *Sea μ una medida de probabilidad en A y ν una medida de probabilidad en B , y sea $\mu \times \nu$ la medida producto en $A \times B$. Si $f : A \times B \rightarrow \mathbb{R}$ es medible con respecto a $\mu \times \nu$, entonces*

$$\int_{A \times B} f d(\mu \times \nu) = \int_A \left(\int_B f(x, y) \nu(dy) \right) \mu(dx) = \int_B \left(\int_A f(x, y) \mu(dx) \right) \nu(dy)$$

con la condición de que $\int_{A \times B} |f| d(\mu \times \nu) < \infty$.

Demostración (ver [28, págs. 110-111])

Veamos ahora un ejemplo que nos muestra que la hipótesis en el teorema anterior es necesaria.

Ejemplo 1.3.1 *Consideremos la sucesión $\{a_{m,n}\}$ con $m, n \in \mathbb{N}$ dada por*

$$a_{m,n} = \begin{cases} 1 & \text{si } m = n \\ -1 & \text{si } m = n + 1 \\ 0 & \text{otro caso} \end{cases}$$

Puede verificarse fácilmente que

$$\sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{m,n} = 0$$

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{m,n} = 1$$

Para ello nos valemos de la siguiente matriz que representa a la sucesión $\{a_{m,n}\}$ anterior:

$$\begin{bmatrix} 1 & -1 & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & -1 & 0 & \cdots & \cdots & \cdots \\ 0 & 0 & 1 & -1 & 0 & \cdots & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \ddots \end{bmatrix}$$

La primera suma corresponde a sumar renglón por renglón y la segunda suma corresponde a sumar columna por columna. Por tanto, el teorema de Fubini no es aplicable puesto que $\{|a_{m,n}|\}$ no es integrable en el espacio $\mathbb{N} \times \mathbb{N}$, con la medida producto μ_C (la medida de contar).

Definición 1.3.1 Sea f una función integrable sobre un intervalo $[a, b]$. La *media de f sobre $[a, b]$* se define como

$$\bar{f} = \frac{1}{b-a} \int_a^b f$$

Teorema 1.3.2 Si f es una función continua sobre $[a, b]$, entonces existe un número real $c \in (a, b)$, tal que $f(c) = \bar{f}$, es decir, tal que

$$\int_a^b f = f(c)(a-b)$$

Demostración (ver [12, págs. 570-571])

1.4. La distribución multinomial

Está asociada con las pruebas repetidas de un suceso con más de dos resultados posibles ([26]). Así, el resultado del lanzamiento de un dado puede ser cualquiera de los seis números 1, 2, 3, ..., 6. Si el suceso se refiere a la aparición de ases en la extracción de, por ejemplo, 7 cartas, hay cinco resultados posibles: 0, 1, 2, 3 o 4 ases. En general, supongamos que hay k resultados posibles de un suceso aleatorio y designemos las probabilidades respectivas de estos resultados por p_1, p_2, \dots, p_k . Evidentemente, debe verificarse que

$$\sum_{i=1}^k p_i = 1$$

Supongamos que el suceso se repite n veces y sea X_1 la variable aleatoria que representa el número de veces que ocurre el resultado asociado a p_1 , X_2 la variable aleatoria que representa el número de que veces que ocurre el resultado asociado a p_2 , y así sucesivamente. La función de distribución para las variables aleatorias X_1, X_2, \dots, X_k es

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k p_i^{x_i}; \quad x_i = 0, 1, \dots, n; \quad \sum_{i=1}^k x_i = n$$

Si definimos el vector aleatorio $X = (X_1, X_2, \dots, X_k)$ y el vector de parámetros $p = (p_1, p_2, \dots, p_k)$, entonces la distribución de X es una **multinomial** de parámetros n y p , y se representa como

$$X \sim \text{Multinomial}(n, p)$$

1.5. Multiplicadores de Lagrange

Este método nos sirve para resolver problemas del tipo general de programación no lineal de la forma

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeto a} && g_1(x) = 0 \quad \text{donde} \quad f : \mathbb{R}^n \rightarrow \mathbb{R} \\ & && g_2(x) = 0 \quad \quad \quad g : \mathbb{R}^n \rightarrow \mathbb{R}^m \\ & && \vdots \\ & && g_m(x) = 0 \end{aligned}$$

con $m \leq n$, donde $x = (x_1, x_2, \dots, x_n)$ y g_i para $i = 1, 2, \dots, m$ son continuas, y en general se supone que tienen segundas derivadas parciales continuas. Por simplicidad notacional, hemos introducido la función con valor vectorial $g = (g_1, g_2, \dots, g_m)$ y volvemos a escribir el problema anterior como

$$\begin{aligned} & \text{minimizar} && f(x) \\ & \text{sujeto a} && g(x) = 0 \end{aligned}$$

Definición 1.5.1 *Un punto x^* que satisfaga la restricción $g(x^*) = 0$ se denomina **punto regular** de la restricción si los vectores gradientes $\nabla g_1(x^*), \nabla g_2(x^*), \dots, \nabla g_m(x^*)$ son linealmente independientes.*

Teorema 1.5.1 *Sea x^* un punto extremo local de f sujeto a las restricciones $g(x) = 0$. Supongamos, además que x^* es un punto regular de estas restricciones. Entonces, existe un $\lambda \in \mathbb{R}^m$ tal que*

$$\nabla f(x^*) + \lambda^T \nabla g(x^*) = 0$$

Demostración (ver [20, pág. 305])

Observamos que las condiciones necesarias de primer orden

$$\nabla f(x^*) + \lambda^T \nabla g(x^*) = 0$$

junto con las restricciones

$$g(x^*) = 0$$

proporcionan un total de $n + m$ ecuaciones (en general, no lineales) en las $n + m$ variables que comprenden (x^*, λ) . Así, las condiciones necesarias son un conjunto completo, pues, al

menos localmente, determinan una solución única. Introducimos el *lagrangiano* asociado al problema con restricciones, definido como

$$l^*(x, \lambda) = f(x) + \lambda^T g(x)$$

Entonces, las condiciones necesarias se pueden expresar de las formas

$$\nabla_x l^*(x, \lambda) = 0$$

$$\nabla_\lambda l^*(x, \lambda) = 0$$

la segunda no es más que un nuevo planteamiento de las restricciones.

Ejemplo 1.5.1 *Consideremos el problema*

$$\begin{aligned} \text{minimizar} \quad & x_1x_2 + x_2x_3 + x_1x_3 \\ \text{sujeto a} \quad & x_1 + x_2 + x_3 = 3 \end{aligned}$$

Identificamos a $f(x)$ y $g(x)$ como

$$f(x) = x_1x_2 + x_2x_3 + x_1x_3 \quad \text{y} \quad g(x) = x_1 + x_2 + x_3 - 3$$

Luego, como tenemos solo una restricción, el lagrangiano asociado al problema sería

$$l^*(x, \lambda) = f(x) + \lambda g(x)$$

$$l^*(x, \lambda) = x_1x_2 + x_2x_3 + x_1x_3 + \lambda(x_1 + x_2 + x_3 - 3)$$

De las condiciones necesarias tendríamos

$$\nabla_x l^*(x, \lambda) = (x_2 + x_3 + \lambda, x_1 + x_3 + \lambda, x_1 + x_2 + \lambda) = (0, 0, 0)$$

$$\nabla_\lambda l^*(x, \lambda) = x_1 + x_2 + x_3 - 3 = 0$$

Se formaría un sistema de cuatro ecuaciones con cuatro incógnitas, así

$$x_2 + x_3 + \lambda = 0$$

$$x_1 + x_3 + \lambda = 0$$

$$x_1 + x_2 + \lambda = 0$$

$$x_1 + x_2 + x_3 - 3 = 0$$

Resolviendo las ecuaciones, obtenemos la solución

$$x_1 = x_2 = x_3 = 1, \quad \lambda = -2$$

Capítulo 2

Teoría de Cadenas de Markov

Tomando como referencia a Hoel ([13]), Karlin ([16]) y Rincón ([27]) vamos a analizar a las cadenas de Markov desde el punto de vista probabilístico. Empezaremos dando algunos ejemplos de procesos estocásticos, siendo las cadenas de Markov un caso particular de estos. Definiremos lo que es una cadena de Markov junto con la propiedad importante de esta conocida como propiedad de Markov, luego daremos otras propiedades que se derivan de la matriz de probabilidades de transición que es una matriz que contiene las probabilidades de ir de un estado a otro. Con la clasificación de los estados de una cadena de Markov podremos analizar el comportamiento límite del sistema.

2.1. Conceptos básicos

Sea $(\Omega, \mathcal{F}, \mathcal{P})$ un espacio de probabilidad y T un conjunto no vacío, finito o numerable. Consideremos una sucesión de variables aleatorias $X_n : \Omega \rightarrow T$. Esto es, para cada $n \in \mathbb{N}$ y $\omega \in \Omega$, tenemos que $X_n(\omega)$ es un elemento del conjunto T . Es usual decir que “el proceso se encuentra en el estado i al tiempo n ” cuando $X_n(\omega) = i$. El conjunto T es el espacio de estados del proceso y en la mayoría de situaciones se trabajará con $T = \{0, 1, 2, \dots\}$.

Definición 2.1.1 *Sea $(\Omega, \mathcal{F}, \mathcal{P})$ un espacio de probabilidad y T un conjunto no vacío. Un proceso estocástico es una colección de variables aleatorias, $X_n : \Omega \rightarrow T$ indexadas por algún conjunto T . En general*

$$T = \begin{cases} \{0, \pm 1, \pm 2, \dots\} & \text{proceso estocástico a tiempo discreto} \\ \mathbb{R} \text{ o } \mathbb{R}^+ & \text{proceso estocástico a tiempo continuo} \end{cases}$$

El conjunto T es el espacio de estados. Si T es numerable se dice que el proceso tiene espacio de estados discreto mientras que si T es continuo se dice que tiene espacio de estados continuo.

Vamos a ver ahora un proceso estocástico muy conocido e importante, y que, como se verá más adelante, es también una cadena de Markov.

2.2. Caminata aleatoria simple

Se le puede considerar como un modelo de juego repetido. Supongamos que se empieza un juego con a soles y se hace sucesivas apuestas de 1 sol. En cada apuesta se tiene una probabilidad p de ganar 1 sol y una probabilidad $1 - p$ de perder 1 sol. Si X_n es la cantidad de dinero que se tiene en el tiempo n (a la que nos referiremos como la fortuna a tiempo n), entonces $X_0 = a$; luego, X_1 puede ser $a + 1$ o $a - 1$ según gane o pierda la primera apuesta; X_2 puede ser $a + 2$, si ha ganado las dos apuestas, a , si ha ganado una y ha perdido la otra o $a - 2$, si ha perdido las dos apuestas. Continuando este procedimiento, se obtendrá una secuencia completa X_0, X_1, X_2, \dots de valores aleatorios que corresponden a su *fortuna* a tiempo $0, 1, 2, \dots$

El proceso estocástico anterior $\{X_n\}_{n \in \mathbb{N}}$ se denomina *caminata aleatoria simple*. También se puede definir este modelo partiendo de $\{Z_i\}$ variables aleatorias independientes e idénticamente distribuidas con $P(Z_i = 1) = p$ y $P(Z_i = -1) = 1 - p$, siendo $0 < p < 1$. Para el caso anterior $Z_i = 1$ si gana la i -ésima apuesta mientras que $Z_i = -1$ si pierde la i -ésima apuesta. Seguidamente establecemos $X_0 = a$ y para $n \geq 1$ hacemos

$$X_n = a + Z_1 + Z_2 + \dots + Z_n$$

Ejemplo 2.2.1 Consideremos el paseo aleatorio simple con $a = 8$ y $p = 1/3$, de forma que se inicia el juego con 8 soles y se tiene una probabilidad de ganar igual a $1/3$ en cada apuesta. La probabilidad de tener 9 soles tras la primera apuesta sería

$$P(X_1 = 9) = P(8 + Z_1 = 9) = P(Z_1 = 1) = 1/3$$

Asimismo, la probabilidad de que se tenga 7 soles después de la primera apuesta es

$$P(X_1 = 7) = P(8 + Z_1 = 7) = P(Z_1 = -1) = 2/3$$

Por otra parte, la probabilidad de que se tenga 10 soles tras las dos primeras apuestas sería

$$P(X_2 = 10) = P(8 + Z_1 + Z_2 = 10) = P(Z_1 + Z_2 = 2) = P(Z_1 = Z_2 = 1) = (1/3)(1/3) = 1/9$$

Ejemplo 2.2.2 Consideremos ahora el paseo aleatorio simple con $a = 8$ y $p = 1/3$. La probabilidad que se tenga 7 soles tras las tres primeras apuestas sería

$$P(X_3 = 7) = P(8 + Z_1 + Z_2 + Z_3 = 7) = P(Z_1 + Z_2 + Z_3 = -1)$$

Ahora, tenemos tres formas diferentes de que $Z_1 + Z_2 + Z_3 = -1$, que pudieran ser

- $Z_1 = 1$ y $Z_2 = Z_3 = -1$
- $Z_2 = 1$ y $Z_1 = Z_3 = -1$
- $Z_3 = 1$ y $Z_1 = Z_2 = -1$

Cada una de las posibilidades anteriores tiene una probabilidad igual a $(1/3)(2/3)(2/3)$. Por tanto

$$P(X_3 = 7) = (1/3)(2/3)(2/3) + (1/3)(2/3)(2/3) + (1/3)(2/3)(2/3) = 4/9$$

Si el número de apuestas es mucho mayor a que tres, esta forma de calcular las probabilidades resulta cada vez menos adecuada y requerimos de un procedimiento más sistemático como el que veremos a continuación.

2.2.1. La distribución de la fortuna

Primero determinaremos la distribución de X_n , es decir, la probabilidad de que la fortuna X_n , tras n apuestas, tome diferentes valores.

Teorema 2.2.1 Sea $\{X_n\}_{n \in \mathbb{N}}$ una caminata aleatoria simple. Si k es un entero tal que $-n \leq k \leq n$ y $n+k$ es un número par, entonces

$$P(X_n = a + k) = \binom{n}{\frac{n+k}{2}} p^{(n+k)/2} q^{(n-k)/2}$$

Para los demás valores de k tenemos $P(X_n = a + k) = 0$. Además, $E(X_n) = a + n(2p - 1)$.

Demostración (ver [8, pág. 705])

Corolario 2.2.1 Sea $\{X_n\}_{n \in \mathbb{N}}$ una caminata aleatoria simple. Si $p = 1/2$, $E(X_n) = a$ para todo $n \geq 0$. Si $p < 1/2$, $E(X_n) < a$ para todo $n \geq 1$. Si $p > 1/2$, $E(X_n) > a$ para todo $n \geq 1$.

Demostración

Sea $n_0 \in \mathbb{N}$ fijo. Del teorema 2.2.1 sabemos que

$$E(X_n) = a + n(2p - 1)$$

Primero, si $p = \frac{1}{2}$, entonces

$$E(X_n) = a + n \left(2 \cdot \frac{1}{2} - 1 \right) .$$

$$E(X_n) = 0$$

Ahora, si $p < \frac{1}{2}$, entonces

$$a + n(2p - 1) < a$$

$$E(X_n) < a$$

Por último, si $p > \frac{1}{2}$, entonces

$$a + n(2p - 1) > a$$

$$E(X_n) > a$$

□

Podemos interpretar el corolario anterior de la siguiente manera: Si $p = 1/2$, el juego es *equilibrado*, es decir, el jugador y su oponente tienen la misma posibilidad de ganar. Por otra parte, si $p < 1/2$, el juego es *subequilibrado*, es decir, las posibilidades de que el oponente gane son mayores que las que gane el jugador. Por último, si $p > 1/2$, el juego es *sobreequilibrado*, es decir, el jugador tendrá más posibilidades de ganar que su oponente.

2.2.2. El problema de la ruina del jugador

Vamos a considerar ahora la evolución en el tiempo del proceso, es decir, considerar diferentes valores de n a la vez.

Sea $\{X_n\}_{n \in \mathbb{N}}$ una caminata aleatoria como el caso anterior, con un capital inicial a y algún valor p de la probabilidad de ganar la cada apuesta, y supongamos que a es un entero positivo. Además, sea $c > a$ otro valor entero. La pregunta sobre este problema es: Si un jugador

apuesta repetidamente 1 sol, ¿cuál es la probabilidad de que disponga de una fortuna de c soles antes de perder todo su dinero, es decir, antes de que su fortuna sea 0 soles? O en forma equivalente, ¿alcanzará la caminata aleatoria el valor c antes de llegar al valor 0?

Formalmente, sean

$$\tau_0 = \min\{n \geq 0 : X_n = 0\}$$

$$\tau_c = \min\{n \geq 0 : X_n = c\}$$

los tiempos en que se llega por *primera vez* a 0 y c , respectivamente, es decir, τ_0 es el tiempo en que se llega por primera vez a un capital 0 mientras que τ_c es la primera vez que la fortuna del jugador llega a ser c .

La pregunta de la ruina del jugador es, ¿cuál es la probabilidad de alcanzar c antes de alcanzar 0?, es decir

$$P(\tau_c < \tau_0)$$

Puesto que no existe un límite respecto a la duración del juego, no es fácil responder a la pregunta anterior, pero podemos solucionarlo de la siguiente forma:

Teorema 2.2.2 *Sea $\{X_n\}_{n \in \mathbb{N}}$ una caminata aleatoria simple como en el caso anterior, con un capital inicial a y algún valor p de la probabilidad de ganar cada apuesta. Supongamos que $0 < a < c$. La probabilidad $P(\tau_c < \tau_0)$ de alcanzar c antes de 0 se define como*

$$P(\tau_c < \tau_0) = \begin{cases} \frac{a}{c}, & p = \frac{1}{2} \\ \frac{1 - \left(\frac{q}{p}\right)^a}{1 - \left(\frac{q}{p}\right)^c}, & p \neq \frac{1}{2} \end{cases}$$

Demostración (ver [8, pág. 705-708])

También podemos utilizar el problema de la ruina del jugador para calcular $P(\tau_0 < \infty)$, la probabilidad de que la caminata alcance siempre 0 o, equivalentemente, de que el jugador siempre pierda toda su dinero, de la siguiente forma:

Teorema 2.2.3 *Sea $\{X_n\}_{n \in \mathbb{N}}$ una caminata aleatoria simple, con un capital inicial $a > 0$ y una probabilidad p de ganar cada apuesta. La probabilidad $P(\tau_0 < \infty)$ de que la caminata aleatoria alcance siempre el valor 0 se define como*

$$P(\tau_0 < \infty) = \begin{cases} 1, & p \leq \frac{1}{2} \\ \left(\frac{q}{p}\right)^a, & p > \frac{1}{2} \end{cases}$$

Demostración (ver [8, pág. 708])

2.3. Cadenas de Markov

Definición 2.3.1 Sea $(\Omega, \mathcal{F}, \mathcal{P})$ un espacio de probabilidad y T un conjunto no vacío, finito o numerable. Un proceso estocástico o sucesión de variables aleatorias $\{X_n : \Omega \rightarrow T, n \in \mathbb{N}\}$ (o simplemente $\{X_n\}_{n \in \mathbb{N}}$) se llama **cadena de Markov** con espacios de estados T si satisface la condición de Markov, es decir, si para todo $n \geq 0$ y toda sucesión $i_0, i_1, \dots, i_{n-1}, i_n, i_{n+1} \in T$ se cumple que

$$P(X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = i_{n+1} | X_n = i_n)$$

Ahora vamos a demostrar algunas propiedades para las cadenas de Markov.

Propiedad 2.3.1 Si $\{X_n\}_{n \in \mathbb{N}}$ es una cadena de Markov, entonces se cumple que

$$P(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_n = i_n | X_{n-1} = i_{n-1}) \dots P(X_1 = i_1 | X_0 = i_0) \cdot P(X_0 = i_0)$$

Demostración

Por inducción, probemos para $n = 1$. Por definición de probabilidad condicional

$$P(X_1 = i_1, X_0 = i_0) = P(X_1 = i_1 | X_0 = i_0) \cdot P(X_0 = i_0)$$

Supongamos que se cumple para m , probemos para $m + 1$. Nuestra hipótesis inductiva sería

$$P(X_m = i_m, X_{m-1} = i_{m-1}, \dots, X_0 = i_0) = \\ P(X_m = i_m | X_{m-1} = i_{m-1}) \dots P(X_1 = i_1 | X_0 = i_0) \cdot P(X_0 = i_0)$$

Luego, para $m + 1$, por definición de probabilidad condicional

$$P(X_{m+1} = i_{m+1}, X_m = i_m, \dots, X_0 = i_0) = \\ P(X_{m+1} = i_{m+1} | X_m = i_m, X_{m-1} = i_{m-1}, \dots, X_0 = i_0) \cdot P(X_m = i_m, X_{m-1} = i_{m-1}, \dots, X_0 = i_0)$$

Por ser una cadena de Markov

$$P(X_{m+1} = i_{m+1}, X_m = i_m, \dots, X_0 = i_0) = \\ P(X_{m+1} = i_{m+1} | X_m = i_m) \cdot P(X_m = i_m, X_{m-1} = i_{m-1}, \dots, X_0 = i_0)$$

Ahora, de la hipótesis inductiva

$$P(X_{m+1} = i_{m+1}, X_m = i_m, \dots, X_0 = i_0) = \\ P(X_{m+1} = i_{m+1} | X_m = i_m) \cdot P(X_m = i_m | X_{m-1} = i_{m-1}) \cdot \dots \cdot P(X_1 = i_1 | X_0 = i_0) \cdot P(X_0 = i_0)$$

Por lo tanto, para todo $n \in \mathbb{N}$ se cumple que

$$P(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_n = i_n | X_{n-1} = i_{n-1}) \cdot \dots \cdot P(X_1 = i_1 | X_0 = i_0) \cdot P(X_0 = i_0)$$

□

Ejemplo 2.3.1 Como lo indicamos anteriormente, vamos a probar que la caminata aleatoria simple $\{X_n\}_{n \in \mathbb{N}}$ es una cadena de Markov. Para ello probaremos el caso general.

Sean Z_1, Z_2, \dots, Z_n variables aleatorias independientes con densidad común $P(Z_n = i) = c_i$ y espacio de estados $T = \{0, \pm 1, \pm 2, \dots\}$. Sea X_0 una variable aleatoria independiente de las variables Z_1, Z_2, \dots, Z_n . Consideremos también

$$X_n = X_0 + Z_1 + \dots + Z_n, \quad n \in \mathbb{N}$$

Con estas definiciones $\{X_n\}_{n \in \mathbb{N}}$ viene a ser una **caminata aleatoria**.

Probaremos que $\{X_n\}_{n \in \mathbb{N}}$ es una cadena de Markov con probabilidades de transición $p_{ij} = c_{j-i}$ con $i, j \in T$.

Como $X_{n+1} = X_n + Z_{n+1}$ y por la independencia de $\{X_k\}_{k=0}^n$ y Z_{n+1} tenemos que

$$P(X_{n+1} = j | X_0 = i_0, \dots, X_n = i) = P(Z_{n+1} = j - X_n | X_0 = i_0, \dots, X_n = i)$$

$$P(X_{n+1} = j | X_0 = i_0, \dots, X_n = i) = P(Z_{n+1} = j - i | X_0 = i_0, \dots, X_n = i)$$

$$P(X_{n+1} = j | X_0 = i_0, \dots, X_n = i) = \frac{P(Z_{n+1} = j - i) \cdot P(X_0 = i_0, \dots, X_n = i)}{P(X_0 = i_0, \dots, X_n = i)}$$

$$P(X_{n+1} = j | X_0 = i_0, \dots, X_n = i) = P(Z_{n+1} = j - i) = c_{j-i}$$

$$P(X_{n+1} = j | X_0 = i_0, \dots, X_n = i) = P(X_{n+1} = j | X_n = i)$$

Por lo tanto

$$p_{ij} = c_{j-i}, \quad i, j \in T$$

Sin pérdida de generalidad tomaremos como espacio de estados de una cadena de Markov al conjunto discreto $\{0, 1, 2, \dots\}$ o cualquier subconjunto finito que conste de los primeros elementos de este conjunto. Cuando el espacio de estados de una cadena de Markov es un conjunto finito se dice que es una *cadena finita*.

A la probabilidad $P(X_{n+1} = j | X_n = i)$ la denotaremos por $p_{ij}(n+1)$ y representará la probabilidad de transición del estado i en el tiempo n , al estado j en el tiempo $n+1$. Estas probabilidades se conocen como *las probabilidades de transición* en un paso. Cuando los números $p_{ij}(n+1)$ no dependen de n se dice que la cadena es *estacionaria* u *homogénea en el tiempo*. Por simplicidad, en este trabajo, se asumirá tal situación de modo que las probabilidades de transición en un paso se escriben simplemente como p_{ij} . Variando los índices i y j , por ejemplo sobre el conjunto de estados $0, 1, 2, \dots$, se obtiene la *matriz de probabilidades de transición* siguiente:

$$P = \begin{array}{c} \begin{array}{cccc} 0 & 1 & 2 & 3 & \dots \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \left[\begin{array}{ccccc} p_{00} & p_{01} & p_{02} & p_{03} & \dots \\ p_{10} & p_{11} & p_{12} & p_{13} & \dots \\ p_{20} & p_{21} & p_{22} & p_{23} & \dots \\ p_{30} & p_{31} & p_{32} & p_{33} & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array} \right] \end{array}$$

La entrada (i, j) de esta matriz es la probabilidad de transición p_{ij} , es decir, la probabilidad de pasar del estado i al estado j en una unidad de tiempo.

Proposición 2.3.1 [27] La matriz de probabilidades de transición $P = (p_{ij})$ cumple las siguientes propiedades:

a. $p_{ij} \geq 0$, para todo $i, j \in T$

b. $\sum_{j \in T} p_{ij} = 1$, $i \in T$

Demostración

a. Es evidente por el hecho de que los p_{ij} son probabilidades.

b. Para cualquier estado i y cualquier entero $n \geq 0$ se tiene que

$$\sum_{j \in T} p_{ij} = \sum_{j \in T} P(X_{n+1} = j | X_n = i)$$

$$\sum_{j \in T} p_{ij} = P\left(\bigcup_{j \in T} \{X_{n+1} = j\} | X_n = i\right)$$

$$\sum_{j \in T} p_{ij} = P(X_{n+1} \in \{0, 1, \dots\} | X_n = i)$$

$$\sum_{j \in T} p_{ij} = P(X_{n+1} \in \{0, 1, \dots\})$$

$$\sum_{j \in T} p_{ij} = 1$$

□

Esto último significa que a partir de cualquier estado i con probabilidad 1 la cadena pasa necesariamente a algún elemento del espacio de estados al siguiente momento. Toda matriz cuadrada que cumpla estas dos condiciones anteriores se dice que es una *matriz estocástica*. Si además la matriz cumple la condición $\sum_{i \in T} p_{ij} = 1$, es decir, cuando la suma de columnas también es 1, entonces se dice que es *doblemente estocástica*.

Ejemplo 2.3.2 Sea $T = \{1, 2, 3\}$ formado por tres elementos, y las probabilidades de transición definidas por $p_{11} = 0$, $p_{12} = 1/2$, $p_{13} = 1/2$, $p_{21} = 1/3$, $p_{22} = 1/3$, $p_{23} = 1/3$, $p_{31} = 1/4$, $p_{32} = 1/4$ y $p_{33} = 1/2$. Esto significa que, si la cadena está en la situación 3, entonces tiene una probabilidad igual a $1/4$ de desplazarse al estado 1 en el siguiente desplazamiento, una probabilidad de $1/4$ de desplazarse al estado 2 en el siguiente desplazamiento y probabilidad de $1/2$ de continuar en el estado 3 en el siguiente movimiento. Esta cadena de Markov se desplaza sobre los tres puntos $\{1, 2, 3\}$ de forma aleatoria. Si se inicia en el punto 1, puede desplazarse al punto 2 o 3 con probabilidad $1/2$ para cada uno de ellos. Si se desplaza a 3, en el siguiente movimiento puede desplazarse a 1 o 2, con probabilidad $1/4$ para cada uno o a 3 con probabilidad igual a $1/2$. Y siempre continuará realizando tales desplazamientos aleatorios.

Podemos expresar las probabilidades de transición p_{ij} en forma matricial, como

$$(p_{ij}) = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

La matriz (p_{ij}) es un ejemplo de matriz estocástica, siendo esta una representación conveniente.

También es necesario establecer el lugar en que empieza la cadena de Markov. Establezcamos sencillamente que $X_0 = t$ para algún estado determinado $t \in T$. De forma general, podríamos tener una distribución inicial $\{\pi_0(i) : i \in T\}$ donde $\pi_0(i) = P(X_0 = i)$. En este caso necesitamos que $\pi_0(i) \geq 0$ para todo $i \in T$, y

$$\sum_{i \in T} \pi_0(i) = 1$$

En resumen, T es el espacio de estados de todos los lugares a los que el objeto puede desplazarse. $\pi_0(i)$ representa la probabilidad de que el objeto parta del punto i , y p_{ij} la probabilidad de que, si el objeto está en el punto i , se desplace al punto j en el siguiente movimiento. En términos de la secuencia de valores aleatorios X_0, X_1, X_2, \dots , tenemos que

$$P(X_{n+1} = j | X_n = i) = p_{ij}$$

para cualquier entero positivo n y cualquier $i, j \in T$. Es necesario también que esta probabilidad de desplazamiento no dependa de la historia previa de la cadena. Es decir, es necesario que

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = p_{ij}$$

para todo n y todo $i, j, i_0, \dots, i_{n-1} \in T$.

La distribución conjunta de las variables aleatorias X_0, X_1, \dots, X_n puede ser expresada en términos de las probabilidades de transición p_{ij} y la distribución inicial π_0 como lo muestra la siguiente proposición:

Proposición 2.3.2 *Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov con $i_0, i_1, \dots, i_n \in T$ y $P(X_0 = i_0) = \pi_0(i_0)$, entonces*

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \pi_0(i_0)p_{i_0i_1} \cdots p_{i_{n-2}i_{n-1}}p_{i_{n-1}i_n} \quad (2.1)$$

Demostración

Usando inducción. Para $n = 1$

$$P(X_0 = i_0, X_1 = i_1) = P(X_0 = i_0) \cdot P(X_1 = i_1 | X_0 = i_0)$$

$$P(X_0 = i_0, X_1 = i_1) = \pi_0(i_0)p_{i_0i_1}$$

Supongamos que se cumple para m

$$P(X_0 = i_0, \dots, X_m = i_m) = \pi_0(i_0)p_{i_0i_1} \cdots p_{i_{m-1}i_m}$$

Probaremos que se cumple para $m + 1$. Usando la definición de probabilidad condicional

$$P(X_0 = i_0, \dots, X_{m+1} = i_{m+1}) = P(X_0 = i_0, \dots, X_m = i_m) \cdot P(X_{m+1} = i_{m+1} | X_0 = i_0, \dots, X_m = i_m)$$

Aplicando la propiedad de Markov en la expresión anterior, tenemos que

$$P(X_0 = i_0, \dots, X_{m+1} = i_{m+1}) = P(X_0 = i_0, \dots, X_m = i_m)p_{i_m i_{m+1}}$$

Pero por la hipótesis inductiva, obtenemos

$$P(X_0 = i_0, X_1 = i_1, \dots, X_{m+1} = i_{m+1}) = \pi_0(i_0)p_{i_0i_1} \cdots p_{i_{m-1}i_m}p_{i_m i_{m+1}}$$

Por lo tanto, para todo $n \in \mathbb{N}$

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \pi_0(i_0)p_{i_0i_1} \cdots p_{i_{n-2}i_{n-1}}p_{i_{n-1}i_n}$$

□

Corolario 2.3.1 Sea $n \in \mathbb{N}$ y $i_0, i_1, \dots, i_n \in T$. Entonces

$$P(X_1 = i_1, \dots, X_n = i_n | X_0 = i_0) = p_{i_0 i_1} \cdot \dots \cdot p_{i_{n-1} i_n}$$

Demostración

De la definición de probabilidad condicional, tenemos que

$$P(X_1 = i_1, \dots, X_n = i_n | X_0 = i_0) = \frac{P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n)}{P(X_0 = i_0)}$$

Notemos que el numerador de la expresión anterior es la distribución conjunta.

Por lo tanto, de la proposición 2.3.2, tenemos que

$$P(X_1 = i_1, \dots, X_n = i_n | X_0 = i_0) = \frac{\pi_0(i_0) p_{i_0 i_1} \cdot \dots \cdot p_{i_{n-2} i_{n-1}} p_{i_{n-1} i_n}}{\pi_0(i_0)}$$

$$P(X_1 = i_1, \dots, X_n = i_n | X_0 = i_0) = p_{i_0 i_1} \cdot \dots \cdot p_{i_{n-2} i_{n-1}} p_{i_{n-1} i_n}$$

□

2.4. Cálculos con cadenas de Markov

Teorema 2.4.1 [8] Considere una cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$ con espacio de estados T , probabilidades de transición p_{ij} y distribución inicial $\pi_0(i)$. Entonces para todo $i \in T$

$$P(X_1 = i) = \sum_{k \in T} \pi_0(k) p_{ki}$$

Demostración

De la ley de probabilidad total

$$P(X_1 = i) = \sum_{k \in T} P(X_0 = k, X_1 = i) \tag{2.2}$$

Pero

$$P(X_0 = k, X_1 = i) = P(X_0 = k) \cdot P(X_1 = i | X_0 = k)$$

$$P(X_0 = k, X_1 = i) = \pi_0(k) p_{ki} \tag{2.3}$$

Reemplazando (2.3) en (2.2)

$$P(X_1 = i) = \sum_{k \in T} \pi_0(k) p_{ki}$$

□

Ejemplo 2.4.1 Sea $T = \{1, 2, 3\}$ y

$$(p_{ij}) = \begin{bmatrix} 1/4 & 1/4 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 0,01 & 0,01 & 0,98 \end{bmatrix}$$

Supongamos que $P(X_0 = 1) = 1/7$, $P(X_0 = 2) = 2/7$ y $P(X_0 = 3) = 4/7$. Entonces

$$P(X_1 = 3) = \sum_{k \in T} \pi_0(k) p_{k3} = (1/7)(1/2) + (2/7)(1/3) + (4/7)(0,98) \approx 0,73$$

Es decir, aproximadamente el 73% de las veces esta cadena se situará en el estado 3 después de una etapa.

Expresemos ahora

$$P_i(A) = P(A|X_0 = i)$$

como la probabilidad del suceso A , suponiendo que la cadena se inicia en el estado i , es decir, suponiendo que $\pi_0(i) = 1$ y $\pi_0(j) = 0$ para $j \neq i$. $P_i(X_0 = j)$ es la probabilidad de que, si la cadena parte del estado i y realiza n etapas, finalizará en el estado j .

Para $n = 0$ tenemos que $X_0 = i$. Por tanto $P_i(X_0 = j) = 1$ si $i = j$, mientras que $P_i(X_0 = j) = 0$ si $i \neq j$. Para $n = 1$, vemos que $P_i(X_1 = j) = p_{ij}$. Es decir, la probabilidad de que estemos en el estado j después de una etapa viene dada por la probabilidad de transición p_{ij} .

Teorema 2.4.2 [8] Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov. Entonces

$$P_i(X_1 = k, X_2 = j) = p_{ik} p_{kj}$$

Demostración

Partiendo de i , la probabilidad de desplazarse en primer lugar a k es p_{ik} . Como en la primera etapa nos hemos desplazado a k , la probabilidad de desplazarnos a j sería p_{kj} .

Por tanto

$$P_i(X_1 = k, X_2 = j) = P(X_1 = k, X_2 = j | X_0 = i)$$

$$P_i(X_1 = k, X_2 = j) = P(X_1 = k | X_0 = i) \cdot P(X_2 = j | X_1 = k, X_0 = i)$$

$$P_i(X_1 = k, X_2 = j) = p_{ik}p_{kj}$$

□

Del teorema 2.4.2 deducimos el siguiente teorema:

Teorema 2.4.3 [8] *Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov. Entonces*

$$P_i(X_2 = j) = \sum_{k \in T} P_i(X_1 = k, X_2 = j)$$

Demostración

De la ley de probabilidad total

$$P_i(X_2 = j) = \sum_{k \in T} P_i(X_1 = k, X_2 = j)$$

Usando el teorema 2.4.2

$$P_i(X_2 = j) = \sum_{k \in T} p_{ik}p_{kj}$$

□

Ejemplo 2.4.2 *Consideremos la siguiente cadena con $T = \{1, 2, 3\}$*

$$(p_{ij}) = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/3 & 1/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

Luego

$$\begin{aligned} P_1(X_2 = 3) &= \sum_{k \in T} p_{1k}p_{k3} \\ &= p_{11}p_{13} + p_{12}p_{23} + p_{13}p_{33} \end{aligned}$$

$$= (0)(1/2) + (1/2)(1/3) + (1/2)(1/2) = 1/6 + 1/4 = 5/12$$

Definición 2.4.1 Dada una cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$ y probabilidades de transición en un paso p_{ij} , se define la **probabilidad de transición de orden n** ($n \in \mathbb{N}$) $p_{ij}^{(n)}$, como la probabilidad de que el proceso vaya del estado i al estado j en n pasos

$$p_{ij}^{(n)} = P(X_{m+n} = j | X_m = i), \quad \forall i, j \in T \quad y \quad m \in \mathbb{N}$$

y

$$p_{ij}^{(0)} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

Es decir, después de realizar cero pasos, la cadena no puede estar en otro estado más que en el origen. La función representada por $p_{ij}^{(0)}$ es la conocida función *delta de Kronecker*. Cuando el espacio de estados T es finito, entonces se denotará $P^{(n)} = (p_{ij}^{(n)})$ como la matriz de probabilidades de transición en n pasos.

Teorema 2.4.4 [16] Las probabilidades de transición de orden n satisfacen que

$$p_{ij}^{(n)} = \sum_{k \in T} p_{ik} p_{kj}^{(n-1)}$$

Demostración

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i)$$

$$p_{ij}^{(n)} = \sum_{k \in T} P(X_n = j, X_1 = k | X_0 = i)$$

$$p_{ij}^{(n)} = \sum_{k \in T} \frac{P(X_n = j, X_1 = k, X_0 = i)}{P(X_0 = i)} \cdot \frac{P(X_1 = k, X_0 = i)}{P(X_1 = k, X_0 = i)}$$

$$p_{ij}^{(n)} = \sum_{k \in T} P(X_1 = k | X_0 = i) \cdot P(X_n = j | X_1 = k, X_0 = i)$$

$$p_{ij}^{(n)} = \sum_{k \in T} p_{ik} p_{kj}^{(n-1)}$$

□

Teorema 2.4.5 [13] Para cualesquiera $m, n \in \mathbb{N}$ y para cualesquiera estados $i, j \in T$ se cumple que

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

Demostración

Tenemos que

$$p_{ij}^{(n+m)} = P(X_{n+m} = j | X_0 = i)$$

$$p_{ij}^{(n+m)} = \frac{P(X_{n+m} = j, X_0 = i)}{P(X_0 = i)}$$

$$p_{ij}^{(n+m)} = P\left(X_{n+m} = j, \bigcup_{k \in T} \{X_n = k\}, X_0 = i\right) \cdot \frac{1}{P(X_0 = i)}$$

$$p_{ij}^{(n+m)} = \sum_{k \in T} \frac{P(X_{n+m} = j, X_n = k, X_0 = i)}{P(X_0 = i)}$$

$$p_{ij}^{(n+m)} = \sum_{k \in T} \frac{P(X_{n+m} = j | X_n = k, X_0 = i) \cdot P(X_n = k, X_0 = i)}{P(X_0 = i)}$$

$$p_{ij}^{(n+m)} = \sum_{k \in T} P(X_{n+m} = j | X_n = k) \cdot P(X_n = k | X_0 = i)$$

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

□

El resultado anterior se conoce como *ecuación de Chapman-Kolmogorov*. Como una consecuencia importante de la ecuación anterior se tiene el siguiente resultado:

Lema 2.4.1 La probabilidad de transición en n pasos, $P^{(n)} = (p_{ij}^{(n)})$, está dada por la entrada (i, j) de la n -ésima potencia de la matriz P , es decir

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

Demostración

Por inducción matemática, para $n = 2$

$$P^{(2)} = (p_{ij}^{(2)})$$

Donde

$$p_{ij}^{(2)} = \sum_{k \in T} p_{ik} p_{kj}, \quad i, j \in T$$

Por definición de producto matricial

$$P^2 = PP = \left(\sum_{k \in T} p_{ik} p_{kj} \right)$$

$$P^2 = (p_{ij}^{(2)})$$

$$P^2 = P^{(2)}$$

Supongamos ahora que se cumple para $n - 1$

$$P^{(n-1)} = (p_{ij}^{(n-1)}) = P^{n-1}$$

Donde

$$p_{ij}^{(n-1)} = \sum_{k \in T} p_{ik} p_{kj}^{(n-2)}$$

Probaremos que es válido para n , es decir,

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

Usando el teorema 2.4.4 tenemos que

$$p_{ij}^{(n)} = \sum_{k \in T} p_{ik} p_{kj}^{(n-1)}$$

Luego

$$P^n = PP^{n-1} = PP^{(n-1)} = (p_{ij})(p_{ij}^{(n-1)}) = \left(\sum_{k \in T} p_{ik} p_{kj}^{(n-1)} \right) = (p_{ij}^{(n)}) = P^{(n)}$$

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

□

Este resultado establece que el problema de calcular las probabilidades de transición en n pasos se transforma en obtener la n -ésima potencia de la matriz de probabilidades de transición en un paso, es decir

$$\begin{bmatrix} p_{00}^{(n)} & p_{01}^{(n)} & \cdots \\ p_{10}^{(n)} & p_{11}^{(n)} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix} = \begin{bmatrix} p_{00} & p_{01} & \cdots \\ p_{10} & p_{11} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}^n$$

El lema 2.4.1 también se puede enunciar en forma equivalente como

$$P_i(X_n = j) = \sum_{i_1, i_2, \dots, i_{n-1} \in T} p_{ii_1} p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{n-2} i_{n-1}} p_{i_{n-1} j}$$

Esto es por el hecho de que, si representamos por P a la matriz (p_{ij}) , por v_0 el vector fila $(\pi_0(i)) = (P(X_0 = i))$ y por v_1 el vector fila $(P(X_1 = i))$, la relación puede expresarse como el producto matricial $v_1 = v_0 P$. Es decir, el vector fila de probabilidades para la cadena después de una etapa v_1 es igual al vector fila de probabilidades para la cadena después de cero etapas v_0 multiplicado por la matriz P de probabilidades de transición. Si representamos por v_n el vector fila $(P(X_n = i))$, por inducción vemos que $v_{n+1} = v_n P$ para cada n , y por tanto, $v_n = v_0 P^n$. De aquí se seguirá entonces que $P_i(X_n = j)$ es igual al elemento (i, j) de la matriz P^n .

Ejemplo 2.4.3 Consideremos nuevamente la cadena del ejemplo 2.3.2 con $T = \{1, 2, 3\}$.

Luego

$$\begin{aligned} P_1(X_3 = 2) &= \sum_{k \in T} \sum_{l \in T} p_{1k} p_{kl} p_{l3} \\ &= p_{11} p_{11} p_{13} + p_{11} p_{12} p_{23} + p_{11} p_{13} p_{33} + \\ &\quad p_{12} p_{21} p_{13} + p_{12} p_{22} p_{23} + p_{12} p_{23} p_{33} + \\ &\quad p_{13} p_{31} p_{13} + p_{13} p_{32} p_{23} + p_{13} p_{33} p_{33} \\ &= (0)(0)(1/2) + (0)(1/2)(1/3) + (0)(1/2)(1/2) + \\ &\quad (1/2)(1/3)(1/2) + (1/2)(1/3)(1/3) + (1/2)(1/4)(1/2) + \end{aligned}$$

$$(1/2)(1/4)(1/2) + (1/2)(1/4)(1/3) + (1/2)(1/2)(1/2) = 31/72 \approx 0,43$$

Es decir, aproximadamente el 43% de las veces esta cadena se situará en el estado 2 después de tres etapas dado que el movimiento de la cadena se inicia en el estado 1.

2.5. Clasificación de estados de una cadena de Markov

A continuación daremos definiciones básicas importantes para el estudio de una cadena de Markov como por ejemplo: comunicación, periodo, primeras visitas, recurrencia, transitoriedad, tiempo medio de recurrencia, número de visitas y la distribución límite de dicha cadena.

2.5.1. Comunicación

Definición 2.5.1 Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov con espacio de estados T y probabilidades de transición p_{ij} , $i, j, \in T$.

- i. Se dice que el estado j es **accesible** desde el estado i si existe un entero $n \geq 0$ tal que $p_{ij}^{(n)} > 0$, lo cual denotaremos como $i \rightarrow j$.
- ii. Se dice además que los estados i y j son **comunicantes** si se cumple que $i \rightarrow j$ y $j \rightarrow i$, lo cual denotaremos como $i \leftrightarrow j$.

Siempre se cumple que $i \leftrightarrow i$, pues por definición $p_{ii}^{(0)} = 1$. También se puede verificar que la comunicación es una relación de equivalencia, es decir, cumple las siguientes propiedades:

- a. Es reflexiva: $i \leftrightarrow i$.
- b. Es simétrica: si $i \leftrightarrow j$, entonces $j \leftrightarrow i$.
- c. Es transitiva: si $i \leftrightarrow j$ y $j \leftrightarrow k$, entonces $i \leftrightarrow k$.

En consecuencia, la comunicación induce una partición del espacio de estados de una cadena de Markov dada por los subconjuntos de estados comunicantes, es decir, dos estados pertenecen al mismo elemento de la partición si y solo si, tales estados se comunican. De este modo el espacio de estados de una cadena de Markov se subdivide en clases de comunicación. A la clase de comunicación de un estado i se le denotará por $C(i)$. Por lo tanto, $i \leftrightarrow j$ si y solo si, $C(i) = C(j)$.

Ejemplo 2.5.1 La cadena de Markov con espacio de estados $T = \{0, 1, 2, 3\}$ y matriz de probabilidades de transición que se muestra en la figura 2.1 tiene tres clases de comunicación que son $C(0) = \{0\}$, $C(1) = \{1, 2\}$ y $C(3) = \{3\}$. Es evidente que el estado 0 se comunica consigo mismo pues existe una flecha que parte de tal estado y llega a él mismo. Visualmente tampoco hay duda de que la segunda colección de estados es una clase de comunicación. Para la clase $C(3)$ no existe una conexión física del estado 3 consigo mismo, sin embargo, por definición $p_{33}^{(0)} = 1$, y ello hace que esta clase de únicamente un estado sea de comunicación. Observe que estas clases de comunicación conforman una partición del espacio de estados.

$$(p_{ij}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$

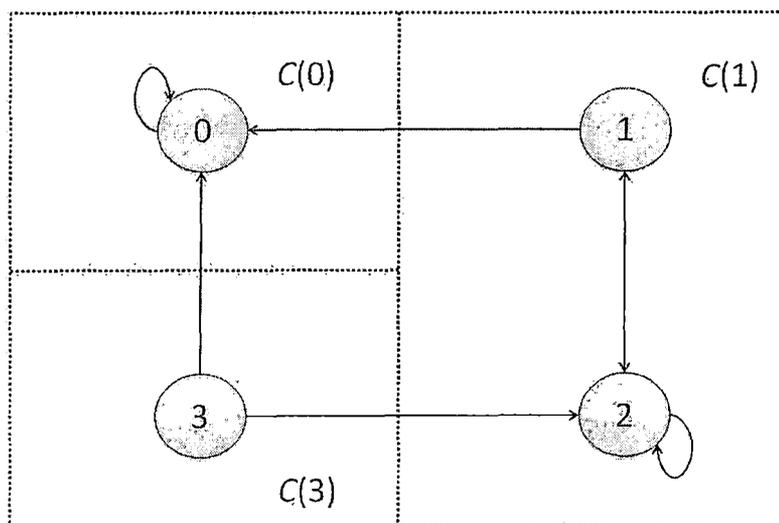


Figura 2.1: Clases de comunicación para el ejemplo 2.5.1

Definición 2.5.2 El estado i de una cadena de Markov se llamará **absorbente** si $p_{ii}^{(1)} = 1$.

Por ejemplo, el estado 0 de la cadena de la figura 2.1 es absorbente.

Definición 2.5.3 Se dice que una cadena de Markov es *irreducible* si todos los estados se comunican entre sí.

Dicho de otra manera, una cadena de Markov es irreducible si existe solo una clase de comunicación, es decir, si la partición generada por la relación de comunicación es trivial. Por ejemplo, la cadena de la caminata aleatoria es una cadena irreducible pues todos los estados se comunican entre sí.

2.5.2. Periodo

El periodo es un número entero no negativo que se calcula para cada estado de una cadena. Se dará interpretación de este número más adelante y aparecerá también dentro de los enunciados generales sobre el comportamiento límite de cadenas de Markov.

Definición 2.5.4 El *periodo* de un estado i es un número entero no negativo denotado por $d(i)$, y definido como sigue

$$d(i) = \text{MCD}\{n \geq 1 : p_{ii}^{(n)} > 0\}$$

Cuando $p_{ii}^{(n)} = 0$ para todo $n \geq 1$, se define $d(i) = 0$. En particular, se dice que un estado i es *aperiódico* si $d(i) = 1$. Cuando $d(i) = k \geq 2$ se dice que i es *periódico* de periodo k .

Proposición 2.5.1 [27] Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov con espacio de estados T y matriz de probabilidades de transición P . Si $i, j \in T$ son comunicantes ($i \longleftrightarrow j$), entonces tienen el mismo periodo ($d(i) = d(j)$).

Demostración

Claramente el resultado es válido para $i = j$. Suponga entonces que i y j son distintos. Como los estados i y j están en la misma clase de comunicación, existen enteros $n \geq 1$ y $m \geq 1$ tales que

$$p_{ij}^{(n)} > 0 \quad \text{y} \quad p_{ji}^{(m)} > 0$$

Sea $s \geq 1$ un entero cualquiera tal que $p_{ii}^{(s)} > 0$. Tal entero existe pues

$$p_{ii}^{(n+m)} \geq p_{ij}^{(n)} p_{ji}^{(m)} > 0$$

Esto quiere decir que $d(i) | s$. Además

$$p_{jj}^{(n+m+s)} \geq p_{ji}^{(m)} p_{ii}^{(s)} p_{ij}^{(n)} > 0$$

Análogamente

$$p_{jj}^{(n+m+2s)} \geq p_{ji}^{(m)} p_{ii}^{(2s)} p_{ij}^{(n)} > 0$$

Por lo tanto $d(j)|(n+m+s)$ y $d(j)|(n+m+2s)$. Entonces $d(j)$ divide a la diferencia $(n+m+2s) - (n+m+s) = s$. Por lo tanto, todo entero $s \geq 1$ tal que $p_{ii}^{(s)} > 0$ cumple $d(j)|s$. Pero $d(i)$ divide a s de manera máxima, por lo tanto

$$d(i) \geq d(j)$$

De manera análoga, escribiendo i por j , y j por i , se obtiene

$$d(j) \geq d(i)$$

Se concluye entonces que

$$d(i) = d(j)$$

□

De lo anterior, vemos que el periodo es una propiedad de clase, es decir, si una cadena de Markov es irreducible, todos sus estados tiene el mismo periodo.

Para probar la siguiente proposición, vamos a utilizar el siguiente lema:

Lema 2.5.1 Sean n_1, \dots, n_k enteros positivos y sea $d = MCD\{n_1, \dots, n_k\}$. Entonces existe un entero positivo M tal que para cada $m \geq M$ existen enteros no negativos c_1, \dots, c_k tales que $md = c_1n_1 + \dots + c_kn_k$.

Demostración (ver [16. pág. 77])

Proposición 2.5.2 [27] Si el estado i tiene periodo $d(i)$, entonces existe un entero N dependiendo de i tal que para todos los enteros $n \geq N$ se cumple que $p_{ii}^{(nd(i))} > 0$.

Demostración

Si $p_{ii}^{(n)} = 0$ para cada $n \geq 1$, entonces $d(i) = 0$ y por lo tanto la afirmación es válida pues $p_{ii}^{(0)} = 1$, sin embargo la interpretación de recurrencia periódica no se aplica en este caso. Suponga entonces que n_1, \dots, n_k son enteros tales que

$$p_{ii}^{(n_1)} > 0 \dots p_{ii}^{(n_k)} > 0$$

Sea $d = MCD\{n_1, \dots, n_k\} \geq d(i)$. Como $d(i)$ es divisor de cada entero n_1, \dots, n_k , se tiene que $d(i)|d$, y por lo tanto existe un entero q tal que $qd(i) = d$. Haciendo uso del lema 2.5.1, entonces tenemos que existe un entero no negativo M tal que para cada $m \geq M$

$$md = c_1n_1 + \dots + c_kn_k$$

para algunos enteros c_1, \dots, c_k , y por lo tanto

$$p_{ii}^{(md)} = p_{ii}^{(c_1n_1 + \dots + c_kn_k)} \geq p_{ii}^{(c_1n_1)} \dots p_{ii}^{(c_kn_k)} > 0$$

Por lo tanto, para cada $m \geq M$

$$p_{ii}^{(md)} = p_{ii}^{(mqd(i))} > 0$$

Definimos $N = Mq$. Se puede entonces concluir que para todo $n \geq N$

$$p_{ii}^{(nd(i))} > 0$$

□

Corolario 2.5.1 [27] Si $p_{ij}^{(m)} > 0$, entonces $p_{ij}^{(m+nd(j))} > 0$ para todo n entero positivo suficientemente grande.

Demostración

A partir de la proposición 2.5.2, para n suficientemente grande, se tiene que

$$p_{ij}^{(m+nd(j))} \geq p_{ij}^{(m)} p_{jj}^{(nd(j))} > 0$$

□

2.5.3. Primeras visitas

También estamos interesados en estudiar el primer momento en el que una cadena de Markov visita un estado particular o un conjunto de estados. Definiremos a continuación este tiempo aleatorio y después demostraremos una fórmula útil que lo relaciona con las probabilidades de transición.

Definición 2.5.5 Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov con espacio de estados T . El tiempo de primera visita al estado i se define como

$$R_i = \begin{cases} \inf \{n \geq 1 : X_n = i\} & \text{si } X_n = i, \text{ para alg\u00fan } n \geq 1 \\ \infty & \text{si } X_n \neq i, \text{ para todo } n > 0 \end{cases}$$

El siguiente resultado relaciona el primer tiempo de entrada con las probabilidades de transici\u00f3n:

Proposici\u00f3n 2.5.3 [13] Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov con espacio de estados T y probabilidades de transici\u00f3n p_{ij} . Entonces se cumple que

$$p_{ij}^{(n)} = \sum_{m=1}^n P(R_j = m | X_0 = i) p_{jj}^{(n-m)} \quad (2.4)$$

con $n \geq 1$ y $i, j \in T$.

Demostraci\u00f3n

Notemos que los eventos $\{R_j = m, X_n = j\}$ con $1 \leq m \leq n$ son disjuntos y adem\u00e1s $\{X_n = j\} = \bigcup_{m=1}^n \{R_j = m, X_n = j\}$. Entonces

$$p_{ij}^{(n)} = \sum_{m=1}^n P(R_j = m, X_n = j | X_0 = i)$$

$$p_{ij}^{(n)} = \sum_{m=1}^n P(R_j = m | X_0 = i) \cdot P(X_n = j | X_0 = i, R_j = m)$$

$$p_{ij}^{(n)} = \sum_{m=1}^n P(R_j = m | X_0 = i) \cdot P(X_n = j | X_0 = i, X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j)$$

$$p_{ij}^{(n)} = \sum_{m=1}^n P(R_j = m | X_0 = i) p_{jj}^{(n-m)}$$

□

Definici\u00f3n 2.5.6 Para cada $n \geq 1$, el n\u00famero $f_{ij}^{(n)}$ denota la probabilidad de que una cadena que inicia en el estado i , llegue al estado j por **primera vez** en exactamente n pasos, es decir

$$f_{ij}^{(n)} = P(R_j = n | X_0 = i)$$

Adicionalmente se define $f_{ij}^{(0)} = 0$, incluyendo el caso $i = j$.

Una forma equivalente de escribir a la probabilidad de primera visita es a través de la expresión

$$f_{ij}^{(n)} = P(X_n = j, X_{n-1} \neq j, \dots, X_1 \neq j | X_0 = i)$$

En particular, observamos que $f_{ii}^{(n)}$ es la probabilidad de regresar por primera vez al mismo estado i en el n -ésimo paso, y que $f_{ii}^{(1)}$ es simplemente p_{ii} . El siguiente resultado establece que la probabilidad de visitar el estado j , a partir de i , en n pasos, puede descomponerse en las probabilidades de los eventos disjuntos en los que se presenta la primera visita, la cual puede efectuarse en el primer paso, o en el segundo paso, y así sucesivamente, hasta el último momento posible n .

De la proposición 2.5.3 y de la definición 2.5.6, vemos que la ecuación (2.4) también sería equivalente a

$$p_{ij}^{(n)} = \sum_{m=1}^n f_{ij}^{(m)} p_{jj}^{(n-m)}$$

2.5.4. Recurrencia y transitoriedad

A continuación veremos que los estados de una cadena de Markov pueden ser clasificados, en una primera instancia, en dos tipos, dependiendo de si la cadena es capaz de regresar con certeza al estado de partida.

Antes de ello vamos a definir, en términos de las probabilidades de primera visita, la probabilidad de una eventual visita al estado j , a partir del estado i .

Definición 2.5.7 Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov con espacio de estados T y matriz de probabilidades de transición P . La probabilidad de que dicha cadena iniciando en i vaya a un estado j en un tiempo positivo se define como

$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)} = \sum_{n=1}^{\infty} P(R_j = n | X_0 = i), \quad i, j \in T$$

El resultado anterior también es equivalente a

$$f_{ij} = P(R_j < \infty | X_0 = i)$$

Definición 2.5.8 Sea $i \in T$. Se define

- i. Un estado i es **recurrente** si $f_{ii} = 1$, es decir, la probabilidad de regresar a él en un tiempo finito es uno.
- ii. Un estado i es **transitorio** si $f_{ii} < 1$, es decir, existe una probabilidad positiva de que la cadena, iniciando en él, ya no regrese nunca a dicho estado.

Antes de enunciar un criterio para determinar si un estado es recurrente o transitorio, definiremos lo que es el número total de visitas y el número esperado de visitas a un cierto estado.

Definición 2.5.9 Sea una cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$ con espacio de estados T . Se define

- i. El **número total de visitas al estado i** como

$$N(i) = \sum_{n=1}^{\infty} \mathbb{I}_{(X_n=i)}$$

donde

$$\mathbb{I}_{(X_n=i)} = \begin{cases} 1, & X_n = i \\ 0, & \text{otro caso} \end{cases}$$

- ii. El **número esperado de visitas a j , dado el estado inicial i** , como

$$G_{ij} = E[N(j)|X_0 = i]$$

Como el evento $\{N(j) \geq 1\}$ es equivalente a $\{R_j < \infty\}$, entonces

$$P(N(j) \geq 1|X_0 = i) = P(R_j < \infty|X_0 = i) = f_{ij}$$

Proposición 2.5.4 (Criterio para la recurrencia) [27] El estado i es

- i. recurrente si y solo si $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.
- ii. transitorio si y solo si $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.

Demostración

Sea $N(i)$ el número total de visitas al estado i , es decir

$$N(i) = \sum_{n=1}^{\infty} \mathbb{I}_{(X_n=i)}$$

Entonces $N(i)$ tiene una distribución geométrica pues para $k \geq 1$

$$P(N(i) \geq k | X_0 = i) = \sum_{m=1}^{\infty} P(N(i) \geq k, X_m = i, X_{m-1} \neq i, \dots, X_1 \neq i | X_0 = i)$$

$$P(N(i) \geq k | X_0 = i) = \sum_{m=1}^{\infty} P(N(i) \geq k | X_m = i, X_{m-1} \neq i, \dots, X_1 \neq i, X_0 = i) \cdot$$

$$P(X_m = i, X_{m-1} \neq i, \dots, X_1 \neq i | X_0 = i)$$

$$P(N(i) \geq k | X_0 = i) = \sum_{m=1}^{\infty} P(N(i) \geq k-1 | X_0 = i) f_{ii}^{(m)}$$

$$P(N(i) \geq k | X_0 = i) = P(N(i) \geq k-1 | X_0 = i) f_{ii}$$

⋮

$$P(N(i) \geq k | X_0 = i) = P(N(i) \geq 1 | X_0 = i) (f_{ii})^{k-1}$$

$$P(N(i) \geq k | X_0 = i) = (f_{ii})^k$$

La esperanza de $N(i)$, posiblemente infinita, puede calcularse de las siguientes dos formas.

Primero

$$E(N(i) | X_0 = i) = \sum_{n=1}^{\infty} P(N(i) \geq n | X_0 = i)$$

$$E(N(i) | X_0 = i) = \sum_{n=1}^{\infty} (f_{ii})^n$$

$$E(N(i) | X_0 = i) = \begin{cases} \frac{f_{ii}}{1-f_{ii}} & \text{si } 0 \leq f_{ii} < 1 \\ \infty & \text{si } f_{ii} = 1 \end{cases} \quad (2.5)$$

Por otro lado, por el teorema de convergencia monótona

$$\begin{aligned}
E(N(i)|X_0 = i) &= E \left[\sum_{n=1}^{\infty} \mathbb{I}_{(X_n=i)} | X_0 = i \right] \\
E(N(i)|X_0 = i) &= \sum_{n=1}^{\infty} E(\mathbb{I}_{(X_n=i)} | X_0 = i) \\
E(N(i)|X_0 = i) &= \sum_{n=1}^{\infty} P(X_n = i | X_0 = i) \\
E(N(i)|X_0 = i) &= \sum_{n=1}^{\infty} p_{ii}^{(n)} \tag{2.6}
\end{aligned}$$

Se concluye la demostración comparando las expresiones anteriores (2.5) y (2.6).

□

Demostraremos a continuación que la recurrencia y la transitoriedad son propiedades de clase, es decir, si dos estados están en una misma clase de comunicación, entonces ambos estados son recurrentes o ambos son transitorios.

Proposición 2.5.5 [27] *La recurrencia es una propiedad de clase, es decir*

- i. Si i es recurrente e $i \longleftrightarrow j$, entonces j es recurrente.
- ii. Si i es transitorio e $i \longleftrightarrow j$, entonces j es transitorio.

Demostración

Como $i \longleftrightarrow j$, existen enteros $n \geq 1$ y $m \geq 1$ tales que $p_{ij}^{(n)} > 0$ y $p_{ji}^{(m)} > 0$. Entonces

$$p_{jj}^{(m+n+r)} \geq p_{ji}^{(m)} p_{ii}^{(r)} p_{ij}^{(n)}$$

De modo que, por la ecuación de Chapman-Kolmogorov

$$\sum_{r=1}^{\infty} p_{jj}^{(m+n+r)} \geq p_{ji}^{(m)} p_{ij}^{(n)} \sum_{r=1}^{\infty} p_{ii}^{(r)}$$

Si i es recurrente, la suma del lado derecho es infinita. Se sigue entonces que la suma del lado izquierdo también lo es, es decir, j es recurrente. La segunda afirmación se demuestra fácilmente por contradicción usando el primer resultado.

□

En consecuencia, cuando una cadena es irreducible y algún estado es recurrente, todos los estados lo son, y se dice que la cadena es recurrente. También puede presentarse la situación en donde el espacio de estados conste de varias clases de comunicación recurrentes, en tal caso la cadena también se llama recurrente. En contraparte, una cadena es transitoria si todos los estados lo son, ya sea conformando una sola clase de comunicación de estados transitorios o varias de ellas. De este modo el espacio de estados de una cadena de Markov puede descomponerse en dos subconjuntos ajenos de estados, aquellos que son transitorios y aquellos que son recurrentes.

A continuación veremos algunos resultados de aplicar los criterios anteriores.

Proposición 2.5.6 [27] *Sea j un estado transitorio. Para cualquier estado inicial i , se cumple que*

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

En consecuencia

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$$

Demostración

Usando la proposición 2.5.6

$$\sum_{i=1}^{\infty} p_{ij}^{(n)} = \sum_{n=1}^{\infty} \sum_{k=1}^n f_{ij}^{(k)} p_{jj}^{(n-k)}$$

La expresión anterior es equivalente a

$$\sum_{i=1}^{\infty} p_{ij}^{(n)} = \sum_{n=1}^{\infty} \sum_{k=0}^{n-1} f_{ij}^{(n-k)} p_{jj}^{(k)}$$

Usando el teorema de Fubini, podemos intercambiar las sumatorias anteriores, puesto que estas son contables

$$\sum_{i=1}^{\infty} p_{ij}^{(n)} = \sum_{k=0}^{\infty} \sum_{n=k+1}^{\infty} f_{ij}^{(n-k)} p_{jj}^{(k)}$$

$$\sum_{i=1}^{\infty} p_{ij}^{(n)} = \sum_{k=0}^{\infty} \sum_{m=1}^{\infty} f_{ij}^{(m)} p_{jj}^{(k)}$$

Luego, como $f_{ij} = \sum_{m=1}^{\infty} f_{ij}^{(m)}$

$$\sum_{i=1}^{\infty} p_{ij}^{(n)} = f_{ij} \sum_{k=0}^{\infty} p_{jj}^{(k)} \leq \sum_{k=0}^{\infty} p_{jj}^{(k)} < \infty$$

$$\sum_{i=1}^{\infty} p_{ij}^{(n)} < \infty$$

□

Definición 2.5.10 Una cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$ es llamada una **cadena de tránsito** si todos sus estados son transitorios y una **cadena recurrente** si todos sus estados son recurrentes.

Proposición 2.5.7 [27] Toda cadena de Markov finita tiene por lo menos un estado recurrente.

Demostración

Por contradicción, supongamos que todos los estados son transitorios. Entonces para cualesquiera estados $i, j \in T$, se cumple que

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

Sumando sobre el conjunto finito de todos los posibles estados j se obtiene

$$\sum_{j \in T} \sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$$

Por otro lado, usando el teorema de Fubini, podemos intercambiar las sumatorias anteriores, puesto que estas son contables, llegándose a la afirmación contraria

$$\sum_{n=1}^{\infty} \sum_{j \in T} p_{ij}^{(n)} = \sum_{i=1}^{\infty} 1 = \infty$$

Por lo tanto es erróneo suponer que todos los estados son transitorios, debe existir por lo menos uno que es recurrente.

□

En consecuencia, toda cadena finita e irreducible es recurrente. Más adelante demostraremos que en tal caso, con probabilidad uno, la cadena visita cada uno de sus estados una infinidad de veces.

Corolario 2.5.2 [13] *Si una cadena de Markov es irreducible, entonces todos los estados son recurrentes o todos son transitorios.*

Demostración

Sea j un estado recurrente. Entonces como la cadena es irreducible los estados i y j son comunicantes ($i \longleftrightarrow j$). Luego por la proposición 2.5.5, i sería un estado recurrente.

Ahora sea j un estado transitorio y supongamos que i es un estado recurrente. Como $i \longleftrightarrow j$, de la proposición 2.5.6, tenemos que j es un estado recurrente lo cual sería una contradicción. Por tanto i necesariamente es un estado transitorio.

□

Corolario 2.5.3 [13] *Si una cadena de Markov es irreducible con un número finito de estados, entonces cada estado es recurrente.*

Demostración

De la proposición 2.5.7 tenemos que existe al menos un estado recurrente. Entonces por el corolario anterior todos los estados son recurrentes.

Sean i y j estados de la cadena, entonces por hipótesis todos los estados se comunican y de la proposición 2.5.2, existe al menos un estado recurrente. Como i se comunica con j , por la proposición 2.5.5, tenemos que j es recurrente. Análogamente se realiza para cada uno de los estados, por lo tanto cada estado es recurrente.

□

2.5.5. Tiempo medio de recurrencia

Hemos visto que si una cadena de Markov inicia en un estado recurrente, entonces regresa a él una infinidad de veces con probabilidad uno. Y hemos definido el tiempo de primera visita al estado i , como la variable aleatoria discreta

$$R_i = \begin{cases} \inf \{n \geq 1 : X_n = i\} & \text{si } X_0 = i, \text{ para algún } n \geq 1 \\ \infty & \text{si } X_0 \neq i, \text{ para todo } n > 0 \end{cases}$$

Vamos a definir el tiempo medio de recurrencia como la esperanza de esta variable aleatoria en el caso cuando el estado a visitar es recurrente.

Definición 2.5.11 *El tiempo medio de recurrencia de un estado recurrente i , se define como*

$$m_i = E[R_i | X_0 = i] = \sum_{n=1}^{\infty} n f_{ii}^n$$

Es importante mencionar que el tiempo medio de recurrencia a un estado puede ser infinito, aun cuando este sea recurrente. A partir de esto, podemos clasificar los estados recurrentes de la siguiente manera:

Definición 2.5.12 *Se dice que un estado recurrente i es*

- i. *recurrente positivo* si $m_i < \infty$.
- ii. *recurrente nulo* si $m_i = \infty$.

Definición 2.5.13 *Una cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$ es llamada*

- i. *cadena recurrente positiva* si cada estado es recurrente positivo.
- ii. *cadena recurrente nula* si cada uno de los estados son recurrentes nulos.

2.5.6. Número de visitas

Los siguientes resultados acerca de la variable aleatoria $N(j)$ y el valor esperado G_{ij} permiten distinguir la diferencia cualitativa en el comportamiento de los estados transitorios respecto de los recurrentes.

Proposición 2.5.8 [13] *Sean $i, j \in T$. Entonces se cumple que*

- i. $P(N(j) = 0 | X_0 = i) = 1 - f_{ij}$
- ii. $P(N(j) \geq m | X_0 = i) = f_{ij}(f_{jj})^{m-1}$
- iii. $P(N(j) = m | X_0 = i) = f_{ij}(f_{jj})^{m-1}(1 - f_{jj})$

Demostración

Para probar i utilizamos la expresión ya vista anteriormente

$$P(N(j) \geq 1 | X_0 = i) = P(R_j < \infty | X_0 = i) = f_{ij}$$

Luego

$$P(N(j) = 0 | X_0 = i) = 1 - P(N(j) \geq 1 | X_0 = i) = 1 - f_{ij}$$

Para probar ii, vamos a denotar la probabilidad de que una cadena de Markov que inició en un estado i , su primera visita al estado j sea en un tiempo m y la siguiente visita j sea en un tiempo n más tarde, con $1 \leq m < n$, como

$$\begin{aligned} & P(X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j, X_{m+1} \neq j, \dots, X_{m+n} = j | X_0 = i) \\ &= P(X_{m+1} \neq j, \dots, X_{m+n+1} \neq j, X_{m+n} = j | X_0 = i, X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j) \\ & \quad \cdot P(X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j | X_0 = i) \\ &= P(X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j | X_0 = j) \cdot P(X_1 \neq j, \dots, X_{m-1} \neq j, X_m = j | X_0 = i) \\ &= P(R_j = m | X_0 = i) \cdot P(R_j = n | X_0 = j) \end{aligned}$$

Por lo tanto

$$\begin{aligned} P(N(j) \geq 2 | X_0 = i) &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} P(R_j = m | X_0 = i) \cdot P(R_j = n | X_0 = j) \\ P(N(j) \geq 2 | X_0 = i) &= \sum_{m=1}^{\infty} P(R_j = m | X_0 = i) \cdot \sum_{n=1}^{\infty} P(R_j = n | X_0 = j) \\ P(N(j) \geq 2 | X_0 = i) &= f_{ij} f_{jj} \end{aligned}$$

Análogamente podemos obtener que la probabilidad de que la cadena de Markov con estado inicial i visite al estado j por primera vez en un tiempo k_1 , la segunda visita sea en un tiempo k_2 después de la primera, la tercera visita sea en un tiempo k_3 después de la segunda, y así sucesivamente la m -ésima visita sea en un tiempo k_m después de la $(m-1)$ -ésima, está dada por

$$P(R_j = k_1 | X_0 = i) \cdot P(R_j = k_2 | X_0 = i) \cdot \dots \cdot P(R_j = k_m | X_0 = i)$$

Entonces

$$P(N(j) \geq m | X_0 = i) = \sum_{k_1=1}^{\infty} \sum_{k_2=1}^{\infty} \dots \sum_{k_m=1}^{\infty} P(R_j = k_1 | X_0 = i) \cdot P(R_j = k_2 | X_0 = j) \cdot \dots \cdot P(R_j = k_m | X_0 = j)$$

$$P(N(j) \geq m | X_0 = i) = \sum_{k_1=1}^{\infty} P(R_j = k_1 | X_0 = i) \cdot \sum_{k_2=1}^{\infty} P(R_j = k_2 | X_0 = j) \cdot \dots \cdot \sum_{k_m=1}^{\infty} P(R_j = k_m | X_0 = j)$$

$$P(N(j) \geq m | X_0 = i) = f_{ij} \underbrace{f_{jj} \cdot \dots \cdot f_{jj}}_{(m-1) \text{ veces}}$$

$$P(N(j) \geq m | X_0 = i) = f_{ij} (f_{jj})^{m-1}$$

Para probar iii, notemos que

$$P(N(j) = m | X_0 = i) = P(N(j) \geq m | X_0 = i) - P(N(j) \geq m + 1 | X_0 = i)$$

$$P(N(j) = m | X_0 = i) = f_{ij} (f_{jj})^{m-1} - f_{ij} (f_{jj})^m$$

$$P(N(j) = m | X_0 = i) = f_{ij} (f_{jj})^{m-1} (1 - f_{jj})$$

□

Proposición 2.5.9 [13] Sea una cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$, con espacio de estados T y probabilidades de transición p_{ij} .

a. Sea j un estado transitorio. Entonces

i. $P(N(j) < \infty | X_0 = i) = 1$, para todo $i \in T$

ii. $G_{ij} = \frac{f_{ij}}{1 - f_{jj}}$, para todo $i \in T$

b. Sea j un estado recurrente. Entonces

i. $P(N(j) = \infty | X_0 = i) = P(R_j < \infty | X_0 = i) = f_{ij}$, para todo $i \in T$

ii. $P(N(j) = \infty | X_0 = j) = 1$ y $G_{jj} = \infty$

iii. Si $f_{ij} = 0$, entonces $G_{ij} = 0$, mientras que si $f_{ij} > 0$, entonces $G_{ij} = \infty$

Demostración

a. Sea j un estado transitorio. Entonces $0 \leq f_{jj} < 1$. Luego usando

$$P(N(j) \geq m | X_0 = i) = f_{ij}(f_{jj})^{m-1}$$

tenemos que

$$P(N(j) = \infty | X_0 = i) = \lim_{m \rightarrow \infty} P(N(j) \geq m | X_0 = i)$$

$$P(N(j) = \infty | X_0 = i) = \lim_{m \rightarrow \infty} f_{ij}(f_{jj})^{m-1}$$

$$P(N(j) = \infty | X_0 = i) = 0$$

lo que implica que

$$P(N(j) < \infty | X_0 = i) = 1$$

De la definición de G_{ij} , tenemos

$$G_{ij} = E[N(j) | X_0 = i]$$

$$G_{ij} = \sum_{m=1}^{\infty} m P(N(j) = m | X_0 = i)$$

Pero

$$P(N(j) = m | X_0 = i) = f_{ij}(f_{jj})^{m-1}(1 - f_{jj})$$

Entonces

$$G_{ij} = \sum_{m=1}^{\infty} m f_{ij}(f_{jj})^{m-1}(1 - f_{jj})$$

$$G_{ij} = f_{ij}(1 - f_{jj}) \sum_{m=1}^{\infty} m (f_{jj})^{m-1}$$

Ahora, haciendo el cambio de variable $q = f_{jj}$ en la serie de potencias

$$\sum_{m=1}^{\infty} m q^{m-1} = \frac{1}{(1 - q)^2}$$

tenemos que

$$G_{ij} = f_{ij}(1 - f_{jj}) \frac{1}{(1 - f_{jj})^2}$$

$$G_{ij} = \frac{f_{ij}}{1 - f_{jj}}$$

Por lo tanto

$$G_{ij} = \frac{f_{ij}}{1 - f_{jj}} < \infty$$

b. Sea j un estado recurrente. Entonces $f_{jj} = 1$ y de

$$P(N(j) \geq m | X_0 = i) = f_{ij}(f_{jj})^{m-1}$$

obtenemos

$$P(N(j) = \infty | X_0 = i) = \lim_{m \rightarrow \infty} P(N(j) \geq m | X_0 = i)$$

$$P(N(j) = \infty | X_0 = i) = \lim_{m \rightarrow \infty} f_{ij}(f_{jj})^{m-1}$$

$$P(N(j) = \infty | X_0 = i) = f_{ij}$$

En particular si $i = j$, $P(N(j) = \infty | X_0 = j) = 1$. Si una variable aleatoria no negativa (como $N(j)$) tiene probabilidad positiva de ser infinita, su esperanza es también infinita. Entonces

$$G_{jj} = E[N(j) | X_0 = j] = \infty$$

Si $f_{ij} = 0$, entonces $P(N(j) = \infty | X_0 = j) = 0$ para todos los enteros positivos m . De la proposición 2.5.3 tenemos que

$$p_{ij}^{(n)} = \sum_{m=1}^n P(R_j = m | X_0 = i) p_{ij}^{(n-m)}, \quad \text{con } n \geq 1 \text{ y } i, j \in T$$

Por lo tanto

$$p_{ij}^{(n)} = \sum_{m=1}^n 0 = 0$$

Usando

$$G_{ij} = E[N(j) | X_0 = i] = \sum_{n=1}^{\infty} p_{ij}^{(n)}$$

tenemos

$$G_{ij} = \sum_{n=1}^{\infty} p_{ij}^{(n)} = 0$$

Ahora, si $f_{ij} > 0$, entonces $P(N(j) = \infty | X_0 = i) = f_{ij} > 0$ y por tanto

$$G_{ij} = E[N(j) | X_0 = i] = \infty$$

□

2.5.7. Distribución límite de una cadena de Markov

En esta sección de la teoría de cadenas de Markov deseamos encontrar, para cada par de estados i y j , condiciones para la existencia de límites de la forma

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi(j)$$

y

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)} = \pi(j)$$

donde π es una distribución.

Para estudiar los límites anteriores, supongamos que tenemos las probabilidades de transición p_{ij} de una cadena de Markov en un espacio de estados T . Sea $\{\pi(i) : i \in T\}$ una distribución definida sobre T tal que $\pi(i) \geq 0$ para todo i y $\sum_{i \in T} \pi(i) = 1$.

Definición 2.5.14 Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov con espacio de estados T y probabilidades de transición p_{ij} , $i, j \in T$

- i. Una distribución de probabilidad π es *estacionaria* o *invariante* para la cadena de Markov si $\pi P = \pi$, es decir

$$\sum_{i \in T} \pi(i) p_{ij} = \pi(j), \quad \text{para todo } j \in T$$

- ii. Una distribución de probabilidad π es una *distribución límite* si

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi(j), \quad \text{para todo } i, j \in T$$

Proposición 2.5.10 [13] Si π es una distribución estacionaria, entonces se cumple que

$$\sum_{i \in T} \pi(i) p_{ij}^{(n)} = \pi(j), \quad \text{para todo } j \in T$$

Demostración

Procediendo por inducción matemática. Sea π una distribución estacionaria. Entonces

$$\begin{aligned} \sum_{i \in T} \pi(i) p_{ij}^{(2)} &= \sum_{i \in T} \pi(i) \sum_{k \in T} p_{ik} p_{kj} \\ \sum_{i \in T} \pi(i) p_{ij}^{(2)} &= \sum_{k \in T} \left(\sum_{i \in T} \pi(i) p_{ik} \right) p_{kj} \\ \sum_{i \in T} \pi(i) p_{ij}^{(2)} &= \sum_{k \in T} \pi(k) p_{kj} \end{aligned}$$

$$\sum_{i \in T} \pi(i) p_{ij}^{(2)} = \pi(j), \quad \text{para todo } j \in T$$

Supongamos que se cumple para m

$$\sum_{i \in T} \pi(i) p_{ij}^{(m)} = \pi(j)$$

Probaremos para $m + 1$

$$\begin{aligned} \sum_{i \in T} \pi(i) p_{ij}^{(m+1)} &= \sum_{i \in T} \pi(i) \sum_{k \in T} p_{ik}^{(m)} p_{kj} \\ \sum_{i \in T} \pi(i) p_{ij}^{(m+1)} &= \sum_{k \in T} \left(\sum_{i \in T} \pi(i) p_{ik}^{(m)} \right) p_{kj} \\ \sum_{i \in T} \pi(i) p_{ij}^{(m+1)} &= \sum_{k \in T} \pi(k) p_{kj} \end{aligned}$$

$$\sum_{i \in T} \pi(i) p_{ij}^{(m+1)} = \pi(j), \quad \text{para todo } j \in T$$

Por lo tanto

$$\sum_{i \in T} \pi(i) p_{ij}^{(n)} = \pi(j), \quad \text{para todo } j \in T$$

□

Se dice que es invariante pues si la distribución inicial π_0 de la cadena es invariante, entonces la distribución de X_n es π_0 para todo $n = 0, 1, \dots$, es decir, π_0 -invariante implica que $P(X_n = j) = \pi_0(j)$ con $j \in T$, para todo $n = 0, 1, \dots$

Proposición 2.5.11 [13] *Si la distribución inicial π_0 es estacionaria, entonces*

$$P(X_n = j) = \pi_0(j), \quad \text{para todo } j \in T$$

Demostración

Sea π_0 una distribución inicial y estacionaria. Entonces

$$\begin{aligned} P(X_n = j) &= \sum_{i \in T} P(X_0 = i, X_n = j) \\ P(X_n = j) &= \sum_{i \in T} P(X_0 = i) \cdot P(X_n = j | X_0 = i) \\ P(X_n = j) &= \sum_{i \in T} \pi_0(i) p_{ij}^{(n)} \\ P(X_n = j) &= \pi_0(j) \end{aligned}$$

□

Ahora vamos a introducir el número total de visitas y el número esperado de visitas de la cadena al tiempo n .

Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov, para cada $n = 0, 1, \dots$. Sea $N_n(j)$ el **número de visitas de la cadena al estado j hasta el tiempo n**

$$N_n(j) = \sum_{m=1}^n \mathbb{I}_{(X_m=j)}$$

Notamos que

$$\lim_{n \rightarrow \infty} N_n(j) = N(j)$$

donde

$$N(j) = \sum_{n=1}^{\infty} \mathbb{I}_{(X_n=j)}$$

Sea G_{ij}^n el número esperado de visitas al estado j hasta el tiempo n , dado que $X_0 = i$

$$G_{ij}^n = E[N_n(j)|X_0 = i] = \sum_{m=1}^n p_{ij}^{(m)}$$

Proposición 2.5.12 [13] *Sea j un estado transitorio. Entonces*

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = 0, \quad \text{con probabilidad 1}$$

y

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = 0, \quad i \in T$$

Demostración

Sea j un estado transitorio. Entonces de la proposición 2.5.9 a.i tenemos que

$$\lim_{n \rightarrow \infty} N_n(j) = N(j) < \infty, \quad \text{con probabilidad 1}$$

Luego

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = 0, \quad \text{con probabilidad 1}$$

De la proposición 2.5.9 a.ii, $G_{ij} < \infty$ para todo $i \in T$, cuando j es transitorio. Entonces

$$\lim_{n \rightarrow \infty} G_{ij}^n = G_{ij} = \sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty, \quad i \in T$$

Por lo tanto

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = 0, \quad i \in T$$

□

Proposición 2.5.13 [13] *Sea j un estado recurrente. Entonces*

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \frac{\mathbb{I}_{(R_j < \infty)}}{m_j}, \quad \text{con probabilidad 1}$$

y

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)} = \lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = \frac{f_{ij}}{m_j}$$

Si $m_j = \infty$, el lado derecho de las igualdades es igual a cero.

Demostración

Consideremos una cadena de Markov que inicia en un estado recurrente j . La cadena regresa a j una infinidad de veces con probabilidad 1. Lo anterior nos garantiza que la variable aleatoria R_j^k , que representa el tiempo de la k -ésima visita al estado j , es finita con probabilidad uno, es decir

$$P(R_j^k < \infty) = 1$$

Definimos las variables aleatorias

$$W_j^1 = R_j^1 \quad \text{y} \quad W_j^k = R_j^k - R_j^{k-1}, \quad \text{para } k \geq 2$$

Notamos que W_j^k representa el tiempo de espera entre la $(k-1)$ y k -ésima visita al estado j .

Entonces se cumple que

$$R_j^k = W_j^1 + \dots + W_j^k$$

Se demuestra que las variables aleatorias W_j^1, W_j^2, \dots son independientes e idénticamente distribuidas (ver [13, pág. 59]), por tanto, tendrán la misma media igual a

$$E[W_j^1 | X_n = j] = E[R_j^1 | X_n = j] = m_j$$

Ahora, usando la ley de los grandes números se tiene que

$$\lim_{n \rightarrow \infty} \frac{W_j^1 + \dots + W_j^n}{n} = m_j, \quad \text{con probabilidad 1}$$

que sería

$$\lim_{n \rightarrow \infty} \frac{R_j^n}{n} = m_j, \quad \text{con probabilidad 1} \quad (2.7)$$

Ahora, sea $k = N_n(j)$ el número de visitas al estado j en las primeras n etapas. Entonces la k -ésima visita a j ocurre antes del tiempo n y la $(k+1)$ -ésima visita a j ocurre después del tiempo n ; esto es

$$R_j^{N_n(j)} \leq n < R_j^{N_n(j)+1}$$

Luego

$$\frac{R_j^{N_n(j)}}{N_n(j)} \leq \frac{n}{N_n(j)} < \frac{R_j^{N_n(j)+1}}{N_n(j)} \quad (2.8)$$

Como j es un estado recurrente se cumple que

$$\lim_{n \rightarrow \infty} N_n(j) = \infty, \quad \text{con probabilidad 1}$$

Entonces de la relación (2.7) y la desigualdad (2.8) tenemos

$$\lim_{n \rightarrow \infty} \frac{n}{N_n(j)} = m_j, \quad \text{con probabilidad 1}$$

que es equivalente a

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \frac{1}{m_j}. \quad \text{con probabilidad 1}$$

Cuando X_0 tiene distribución arbitraria, la cadena puede no alcanzar al estado j , pero si lo alcanza, el argumento anterior también es válido; por tanto

$$\frac{N_n(j)}{n} \xrightarrow{n \rightarrow \infty} \frac{\mathbb{I}_{(R_j < \infty)}}{m_j}, \quad \text{con probabilidad 1} \quad (2.9)$$

Para mostrar la segunda parte, usando $0 \leq N_n(j) \leq n$, tenemos que

$$0 \leq \frac{N_n(j)}{n} \leq 1 \quad (2.10)$$

Por el teorema de convergencia dominada, y a partir de (2.9) y (2.10), concluimos que

$$\lim_{n \rightarrow \infty} E \left[\frac{N_n(j)}{n} \middle| X_0 = i \right] = E \left[\frac{\mathbb{I}_{(R_j < \infty)}}{m_j} \middle| X_0 = i \right]$$

$$\lim_{n \rightarrow \infty} E \left[\frac{N_n(j)}{n} \middle| X_0 = i \right] = \frac{P(R_j < \infty | X_0 = i)}{m_j}$$

$$\lim_{n \rightarrow \infty} E \left[\frac{N_n(j)}{n} \middle| X_0 = i \right] = \frac{f_{ij}}{m_j}$$

Recordemos que

$$G_{ij}^m = E[N_n(j) | X_0 = i] = \sum_{m=1}^n p_{ij}^{(m)}$$

representa el número esperado de visitas al estado j hasta el tiempo n dado $X_0 = i$.

Por tanto

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^m}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)}$$

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^m}{n} = \lim_{n \rightarrow \infty} E \left[\frac{N_n(j)}{n} \middle| X_0 = i \right]$$

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^m}{n} = \frac{f_{ij}}{m_j}$$

□

Corolario 2.5.4 [13] *De la proposición 2.5.13 podemos observar que cuando un estado j es*

i. *recurrente nulo, es decir, $1/m_j = 0$, obtenemos*

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)} = 0, \quad i \in T$$

ii. *recurrente positivo, es decir, $1/m_j > 0$, entonces*

$$\lim_{n \rightarrow \infty} \frac{G_{jj}^n}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{jj}^{(k)} = \frac{1}{m_j} > 0$$

Además, de la proposición 2.5.13 y el corolario 2.5.4, tenemos que el $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)}$ siempre existe y para cada $i \in T$ satisface

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n p_{ij}^{(k)} = \begin{cases} 0, & \text{si } j \text{ es transitorio o recurrente nulo} \\ \frac{1}{m_j}, & \text{si } i = j, j \text{ es recurrente positivo} \\ \frac{f_{ij}}{m_j}, & \text{si } j \text{ es recurrente} \end{cases}$$

Proposición 2.5.14 [13] *Sea i un estado recurrente positivo y suponga que i se comunica con j . Entonces j es un estado recurrente positivo.*

Demostración

De la proposición 2.5.5 se sigue que j se comunica con i , con $j \in T$. Entonces existen enteros positivos n_1 y n_2 tal que

$$p_{ij}^{(n_1)} > 0 \quad \text{y} \quad p_{ji}^{(n_2)} > 0$$

De la ecuación de Chapman-Kolmogorov obtenemos

$$p_{jj}^{(n_1+m+n_2)} \geq p_{ji}^{(n_2)} p_{ii}^{(m)} p_{ij}^{(n_1)}$$

Sumando sobre $m = 1, \dots, n$ y dividiendo por n , tenemos

$$\begin{aligned} \frac{1}{n} \sum_{m=1}^n p_{jj}^{(n_1+m+n_2)} &\geq \frac{1}{n} \sum_{m=1}^n p_{ji}^{(n_2)} p_{ii}^{(m)} p_{ij}^{(n_1)} \\ \sum_{m=1}^{n_1+n+n_2} \frac{p_{jj}^{(m)}}{n} - \sum_{m=1}^{n_1+n_2} \frac{p_{jj}^{(m)}}{n} &\geq p_{ji}^{(n_2)} p_{ij}^{(n_1)} \sum_{m=1}^n \frac{p_{ii}^{(m)}}{n} \\ \frac{G_{jj}^{n_1+n+n_2}}{n} - \frac{G_{jj}^{n_1+n_2}}{n} &\geq p_{ji}^{(n_2)} p_{ij}^{(n_1)} \cdot \frac{1}{n} \sum_{m=1}^n p_{ii}^{(m)} \end{aligned}$$

Pero

$$G_{ii}^n = \sum_{m=1}^n p_{ii}^{(m)}$$

Por tanto

$$\frac{G_{jj}^{n_1+n+n_2}}{n} - \frac{G_{jj}^{n_1+n_2}}{n} \geq p_{ji}^{(n_2)} p_{ij}^{(n_1)} \cdot \frac{G_{ii}^n}{n}$$

Cuando $n \rightarrow \infty$, del corolario 2.5.4 ii tenemos que

$$\lim_{n \rightarrow \infty} \frac{G_{jj}^{n_1+n+n_2}}{n} = \frac{1}{m_j}, \quad \lim_{n \rightarrow \infty} \frac{G_{jj}^{n_1+n_2}}{n} = 0 \quad \text{y} \quad \lim_{n \rightarrow \infty} \frac{G_{ii}^n}{n} = \frac{1}{m_i}$$

Entonces

$$\frac{1}{m_j} \geq \frac{p_{ji}^{(n_2)} p_{ij}^{(n_1)}}{m_i} > 0$$

y consecuentemente $m_j < \infty$. Por tanto j es recurrente positivo. □

Proposición 2.5.15 [13] *Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov con espacio de estados T finito. Entonces se tiene al menos un estado recurrente positivo.*

Demostración

Notemos que para cada $m = 1, 2, \dots$, se cumple que

$$\sum_{j \in T} p_{ij}^{(m)} = 1, \quad i \in T$$

Sumando sobre $m = 1, \dots, n$ y dividiendo por n tenemos que

$$\sum_{m=1}^n \sum_{j \in T} \frac{1}{n} p_{ij}^{(m)} = \sum_{j \in T} \frac{G_{ij}^n}{n} = 1, \quad i \in T$$

Luego

$$\lim_{n \rightarrow \infty} \sum_{j \in T} \frac{G_{ij}^n}{n} = 1, \quad i \in T$$

Ahora, si cada estado fuera transitorio o recurrente nulo, entonces por el corolario 2.5.4 i tenemos que

$$1 = \lim_{n \rightarrow \infty} \sum_{j \in T} \frac{G_{ij}^n}{n} = \sum_{j \in T} \lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = 0, \quad i \in T$$

lo cual sería una contradicción. Por tanto, se tiene al menos un estado recurrente.

□

Corolario 2.5.5 [13] *Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov finita e irreducible. Entonces cada estado es recurrente positivo.*

Demostración

Como la cadena de Markov es finita, esta tiene al menos un estado recurrente positivo. Además, como $\{X_n\}_{n \in \mathbb{N}}$ es irreducible, de la proposición 2.5.14 tenemos que cada estado es recurrente positivo.

□

Corolario 2.5.6 *Una cadena de Markov irreducible con un número finito de estados es recurrente positiva.*

Demostración (ver [13, pág. 62])

Corolario 2.5.7 *Una cadena de Markov con un número finito de estados no tiene estados recurrentes nulos.*

Demostración (ver [13, pág. 62])

Proposición 2.5.16 [13] *Sea π una distribución estacionaria. Si j es un estado transitorio o recurrente nulo, entonces $\pi(j) = 0$.*

Demostración

Si j es un estado transitorio o recurrente nulo, entonces

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = 0, \quad j \in T$$

Sea π una distribución estacionaria. Por la proposición 2.5.10 tenemos que

$$\sum_{i \in T} \pi(i) p_{ij}^{(k)} = \pi(j), \quad k \geq 1 \quad (2.11)$$

Sumando $k = 1, \dots, n$ y dividiendo por n , obtenemos

$$\sum_{i \in T} \pi(i) \frac{G_{ij}^n}{n} = \pi(j), \quad j \in T \quad (2.12)$$

De (2.11), (2.12) y del teorema de convergencia dominada tenemos que

$$\pi(j) = \lim_{n \rightarrow \infty} \sum_{i \in T} \pi(i) \frac{G_{ij}^n}{n} = 0$$

Por lo tanto

$$\pi(j) = 0$$

□

Proposición 2.5.17 [13] *Una cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$ irreducible y recurrente positiva tiene una única distribución estacionaria π , dada por*

$$\pi(j) = \frac{1}{m_j}, \quad j \in T$$

Demostración

De la proposición 2.5.13 y de las hipótesis de esta proposición, concluimos que

$$\lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} = \frac{1}{m_j}, \quad i, j \in T \quad (2.13)$$

Supongamos que π es una distribución estacionaria. De (2.12), (2.13) y del teorema de convergencia dominada tenemos que

$$\pi(j) = \lim_{n \rightarrow \infty} \sum_{i \in T} \pi(i) \frac{G_{ij}^n}{n}$$

$$\pi(j) = \frac{1}{m_j} \sum_{i \in T} \pi(i)$$

$$\pi(j) = \frac{1}{m_j}$$

Entonces, si existe una distribución estacionaria, debe estar dada por

$$\pi(j) = \frac{1}{m_j}, \quad j \in T$$

Para probar la unicidad, supongamos que existe otra distribución estacionaria π' para la cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$. Entonces

$$\pi'(j) = \sum_{i \in T} \pi'(i) p_{ij}^{(m)}, \quad \forall j \in T$$

Sumando sobre $m = 1, \dots, n$ y dividiendo por n a la expresión anterior, tenemos que

$$\begin{aligned} \frac{1}{n} \sum_{m=1}^n \pi'(j) &= \frac{1}{n} \sum_{m=1}^n \sum_{i \in T} \pi'(i) p_{ij}^{(m)} \\ \pi'(j) &= \sum_{i \in T} \pi'(i) \left(\frac{1}{n} \sum_{m=1}^n p_{ij}^{(m)} \right) \end{aligned}$$

Tomando límite cuando $n \rightarrow \infty$ a ambos lados de la expresión anterior

$$\begin{aligned} \lim_{n \rightarrow \infty} \pi'(j) &= \lim_{n \rightarrow \infty} \sum_{i \in T} \pi'(i) \left(\frac{1}{n} \sum_{m=1}^n p_{ij}^{(m)} \right) \\ \pi'(j) &= \sum_{i \in T} \pi'(i) \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{m=1}^n p_{ij}^{(m)} \right) \\ \pi'(j) &= \sum_{i \in T} \pi'(i) \left(\lim_{n \rightarrow \infty} \frac{G_{ij}^n}{n} \right) \end{aligned}$$

Nuevamente usando la ecuación (2.13)

$$\begin{aligned} \pi'(j) &= \sum_{i \in T} \pi'(i) \frac{1}{m_j} \\ \pi'(j) &= \frac{1}{m_j} \sum_{i \in T} \pi'(i) \\ \pi'(j) &= \frac{1}{m_j} \end{aligned}$$

Pero como

$$\pi(j) = \frac{1}{m_j}$$

Por lo tanto

$$\pi(j) = \pi'(j), \quad \forall j \in T$$

Esto prueba la unicidad de la distribución estacionaria π .

Ahora debemos ver que la distribución anterior es estacionaria. Como j es recurrente positivo, tenemos que $m_j < \infty$, entonces $\pi(j) > 0$. Ahora hay que verificar que

$$\sum_{j \in T} \frac{1}{m_j} = 1 \tag{2.14}$$

y que

$$\sum_{j \in T} \frac{1}{m_j} p_{jk} = \frac{1}{m_k}, \quad k \in T \tag{2.15}$$

Observamos que

$$\sum_{j \in T} p_{ij}^{(m)} = 1$$

Sumando $m = 1, \dots, n$ y dividiendo entre n la expresión anterior concluimos que

$$\sum_{j \in T} \frac{G_{ij}^n}{n} = 1, \quad i \in T \tag{2.16}$$

De la ecuación de Chapman-Kolmogorov obtenemos

$$\sum_{j \in T} p_{ij}^{(m)} p_{jk} = p_{ik}^{(m+1)}$$

Otra vez, sumando $m = 1, \dots, n$ y dividiendo por n la expresión anterior, tenemos

$$\sum_{j \in T} \frac{G_{ij}^m}{n} p_{jk} = \frac{G_{ik}^{n+1}}{n} - \frac{p_{ik}}{n} \tag{2.17}$$

Si T es finito, tenemos de (2.13) y (2.16) que

$$1 = \lim_{n \rightarrow \infty} \sum_{j \in T} \frac{G_{ij}^m}{n} = \sum_{j \in T} \frac{1}{m_j}$$

entonces se cumple (2.14).

Ahora tomando el límite a (2.17) cuando $n \rightarrow \infty$ tendríamos

$$\sum_{i \in T} \frac{1}{m_i} p_{ij} = \lim_{n \rightarrow \infty} \sum_{i \in T} \frac{G_{ki}^n}{n} p_{ij}$$

$$\sum_{i \in T} \frac{1}{m_i} p_{ij} = \lim_{n \rightarrow \infty} \frac{G_{kj}^{n+1}}{n} - \lim_{n \rightarrow \infty} \frac{p_{kj}}{n}$$

$$\sum_{i \in T} \frac{1}{m_i} p_{ij} = \frac{1}{m_j}$$

Luego se cumple (2.15). Por tanto, π es una distribución estacionaria cuando el espacio de estados es infinito.

Si T es infinito, la argumentación anterior no es válida, pues en general, no se podría aplicar el teorema de convergencia dominada. Decimos entonces, sea T_1 un subconjunto finito de T . De (2.16) tenemos que

$$\sum_{j \in T_1} \frac{G_{ij}^n}{n} \leq 1, \quad i \in T$$

Como T_1 es finito, cuando $n \rightarrow \infty$ en la desigualdad anterior y por (2.13) tenemos

$$\sum_{j \in T_1} \frac{1}{m_j} \leq 1$$

la cual se cumple para cualquier subconjunto finito $T_1 \subset T$. Por tanto

$$\sum_{j \in T} \frac{1}{m_j} \leq 1 \tag{2.18}$$

pues de lo contrario, la suma sobre algún conjunto finito $T_1 \subset T$ también excederá a 1.

Por un argumento similar, concluimos de (2.18) que si T_1 es un subconjunto finito de T , entonces

$$\sum_{j \in T_1} \frac{G_{ij}^n}{n} p_{jk} \leq \frac{G_{ik}^{n+1}}{n} - \frac{p_{ik}}{n}$$

Tomando límite cuando $n \rightarrow \infty$ en la desigualdad anterior y usando (2.13), obtenemos

$$\sum_{j \in T_1} \frac{1}{m_j} p_{jk} \leq \frac{1}{m_k}$$

Análogamente a la demostración de (2.18), concluimos que

$$\sum_{j \in T} \frac{1}{m_j} p_{jk} \leq \frac{1}{m_k}, \quad k \in T \quad (2.19)$$

Observamos que se cumple la igualdad en la desigualdad anterior. Si no fuera así, entonces de (2.14), y sumando sobre k , se cumple que

$$\begin{aligned} \sum_{k \in T} \frac{1}{m_k} &> \sum_{k \in T} \left(\sum_{j \in T} \frac{1}{m_j} p_{jk} \right) \\ \sum_{k \in T} \frac{1}{m_k} &> \sum_{j \in T} \frac{1}{m_j} \left(\sum_{k \in T} p_{jk} \right) \\ \sum_{k \in T} \frac{1}{m_k} &> \sum_{j \in T} \frac{1}{m_j} \end{aligned}$$

$$1 > 1$$

lo cual sería una contradicción. Por lo tanto, se cumple la igualdad en (2.19), es decir

$$\sum_{j \in T} \frac{1}{m_j} p_{jk} = \frac{1}{m_k}, \quad k \in T \quad (2.20)$$

Por último, debemos demostrar que

$$\sum_{j \in T} \frac{1}{m_j} = 1$$

Definimos

$$a = \frac{1}{\sum_{j \in T} \frac{1}{m_j}}$$

Luego por (2.20)

$$\pi(j) = \frac{a}{m_j}, \quad j \in T$$

define una distribución estacionaria. Entonces de

$$\pi(j) = \frac{1}{m_j}$$

y de la unicidad de la distribución estacionaria π , tenemos que

$$\frac{a}{m_j} = \frac{1}{m_j}$$

Así, concluimos que $a = 1$. Por lo tanto

$$\sum_{j \in T} \frac{1}{m_j} = 1$$

□

Corolario 2.5.8 [13] *Una cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$ irreducible es recurrente positiva si y solo si tiene una distribución estacionaria.*

Demostración

Usando la hipótesis del corolario y de la proposición 2.5.8 tenemos una única distribución estacionaria dada por

$$\pi(j) = \frac{1}{m_j}, \quad j \in T$$

Como π es una distribución estacionaria, entonces si j es un estado transitorio o recurrente nulo, de la proposición 2.5.16, tenemos que $\pi(j) = 0$.

Por lo tanto, cada estado $i \in T$ debe ser recurrente positivo y así la cadena será recurrente positiva.

□

Proposición 2.5.18 (Ley fuerte de los grandes números para cadenas de Markov)
[13] *Si $\{X_n\}_{n \in \mathbb{N}}$ es una cadena de Markov irreducible, recurrente positiva y con distribución estacionaria π , entonces*

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \pi(j), \quad j \in T$$

Demostración

De la proposición 2.5.13 tenemos que

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \frac{1}{m_j}, \quad j \in T$$

Ahora, de la proposición 2.5.17, tenemos que una cadena de Markov irreducible y recurrente positiva tiene una distribución dada por $\pi(j) = 1/m_j$.

Por lo tanto

$$\lim_{n \rightarrow \infty} \frac{N_n(j)}{n} = \pi(j), \quad j \in T$$

□

Corolario 2.5.9 [13] *Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov irreducible. Si el espacio de estados T es finito, entonces tiene una única distribución estacionaria π .*

Demostración

Del corolario 2.5.6, tenemos que si una cadena de Markov es irreducible con espacio de estados finito, entonces la cadena es recurrente positiva. Por lo tanto, de la proposición 2.5.8, se tiene una única distribución estacionaria.

□

Como resumen a lo visto anteriormente, para una cadena de Markov, podemos decir que

- i. si no tiene estados recurrentes positivos, entonces no tiene distribución estacionaria.
- ii. si tiene estados recurrentes positivos y son comunicantes, entonces tiene una única distribución estacionaria.
- iii. si no son comunicantes, entonces tiene infinidad de distribuciones estacionarias.

(ver [13, pág. 68])

Teorema 2.5.1 (Teorema básico de convergencia) [13] *Sea $\{X_n\}_{n \in \mathbb{N}}$ una cadena de Markov irreducible, recurrente positiva y con distribución estacionaria π . Si la cadena es aperiódica, entonces*

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \frac{1}{m_j} = \pi(j), \quad i, j \in T$$

Demostración

Consideremos una cadena de Markov aperiódica, irreducible y recurrente positiva con probabilidades de transición p_{ij} , espacio de estados T y distribución estacionaria π .

Sea $c \in T$. Definimos el conjunto I como

$$I = \{n > 0 : p_{cc}^{(n)} > 0\}$$

Luego

i. $MCD(I) = 1$

ii. Si $m \in I$ y $n \in I$, entonces $(m + n) \in I$.

A partir de la demostración en [13, págs. 79-80], vemos que i y ii implican que existe un entero positivo n' tal que $n \in I$ para todo $n \geq n'$. Luego se cumple que

$$p_{cc}^{(n)} > 0. \quad \text{para } n \geq n'$$

Ahora sean $i, j \in T$. Como la cadena es irreducible, existen enteros positivos n_1 y n_2 tal que

$$p_{ic}^{(n_1)} > 0 \quad \text{y} \quad p_{cj}^{(n_2)} > 0$$

Luego, de la ecuación de Chapman-Kolmogorov para $n \geq n'$ tenemos que

$$p_{ij}^{(n_1+n+n_2)} \geq p_{ic}^{(n_1)} p_{cc}^{(n)} p_{cj}^{(n_2)} > 0$$

Así, para cada $i, j \in T$, existe un entero positivo $n_0 = n_1 + n' + n_2$ tal que

$$p_{ij}^{(n)} > 0, \quad \text{para } n \geq n_0 \tag{2.21}$$

Sea ahora

$$T^2 = T \times T = \{(i, j) : i, j \in T\}$$

el producto cartesiano formado por los pares ordenados de elementos en T . Consideraremos una cadena de Markov (X_n, Y_n) con espacio de estados T^2 y probabilidades de transición definidas por

$$p_2((i_0, j_0), (i, j)) = p_{i_0 i} p_{j_0 j}$$

Tenemos entonces que $\{X_n\}_{n \in \mathbb{N}}$ y $\{Y_n\}_{n \in \mathbb{N}}$ son cadenas de Markov con probabilidades de transición p_{ij} , donde dichas probabilidades de X_n y Y_n son independientes entre sí.

Vamos a demostrar ahora que la cadena de Markov (X_n, Y_n) es aperiódica, irreducible y recurrente positiva.

Sean $(i_0, j_0), (i, j) \in T^2$. Por (2.21) tenemos que existe un n_0 tal que

$$p_{i_0 i}^{(n)} > 0 \quad \text{y} \quad p_{j_0 j}^{(n)} > 0, \quad n \geq n_0$$

Entonces

$$p_2^{(n)}((i_0, j_0), (i, j)) = p_{i_0 i}^{(n)} p_{j_0 j}^{(n)} > 0, \quad n \geq n_0$$

Por tanto, la cadena es aperiódica e irreducible.

Probaremos ahora que la distribución π_2 en T^2 definida como $\pi_2(i_0, j_0) = \pi(i_0)\pi(j_0)$ es estacionaria.

$$\begin{aligned} & \sum_{(i_0, j_0) \in T^2} \pi_2(i_0, j_0) p_2((i_0, j_0), (i, j)) \\ &= \sum_{i_0 \in T} \sum_{j_0 \in T} \pi(i_0) \pi(j_0) p_{i_0 i} p_{j_0 j} \\ &= \left(\sum_{i_0 \in T} \pi(i_0) p_{i_0 i} \right) \left(\sum_{j_0 \in T} \pi(j_0) p_{j_0 j} \right) \\ &= \pi(i) \pi(j) = \pi_2(i, j) \end{aligned}$$

Por tanto, π_2 es una distribución estacionaria para la cadena (X_n, Y_n) . Por el corolario 2.5.8 vemos que esta cadena es recurrente positiva, y en particular, recurrente.

Ahora sea

$$L = \text{mín}\{n > 0 : X_n = Y_n\}$$

y $b \in T$. Como la cadena (X_n, Y_n) es recurrente

$$L_{bb} = \text{mín}\{n > 0 : (X_n, Y_n) = (b, b)\}$$

es finito con probabilidad 1. Además, $L \leq L_{bb}$. Por tanto, L es finito con probabilidad 1. Demostraremos que para algún $n \geq 1$

$$P(X_n = j, L \leq n) = P(Y_n = j, L \leq n), \quad j \in T \quad (2.22)$$

El argumento anterior es razonable, ya que las cadenas son indistinguibles para $n \geq L$.

Para probar (2.22), tomaremos $1 \leq m \leq n$. Entonces para un $k \in T$ tenemos

$$P(X_n = j | L = m, X_m = Y_m = k) = P(Y_n = j | L = m, X_m = Y_m = k) \quad (2.23)$$

ya que las probabilidades condicionales anteriores son iguales a $p_{kj}^{(n-m)}$. Además, el evento $\{L \leq n\}$ es la unión de los eventos disjuntos

$$\{L = m, X_m = Y_m = k\}, \quad 1 \leq m \leq n \quad \text{y} \quad k \in T$$

Luego de (2.23) se sigue que

$$P(X_n = j | L \leq n) = P(Y_n = j | L \leq n), \quad j \in T$$

Usando la igualdad anterior tendríamos que

$$P(X_n = j) = P(X_n = j, L \leq n) + P(X_n = j | L > n)$$

$$P(X_n = j) = P(Y_n = j, L \leq n) + P(X_n = j | L > n)$$

$$P(X_n = j) \leq P(Y_n = j) + P(L > n)$$

De forma análoga podemos obtener

$$P(Y_n = j) \leq P(X_n = j) + P(L > n)$$

Así, para $n \geq 1$ se cumple que

$$|P(X_n = j) - P(Y_n = j)| \leq P(L > n), \quad j \in T \quad (2.24)$$

Por otra parte, como L es finito con probabilidad 1. entonces

$$\lim_{n \rightarrow \infty} P(L > n) = 0 \quad (2.25)$$

De (2.24) y (2.25), tenemos que

$$\lim_{n \rightarrow \infty} (P(X_n = j) - P(Y_n = j)) = 0, \quad j \in T \quad (2.26)$$

Ahora, sea $i \in T$ y sea la distribución inicial de (X_n, Y_n) tal que

$$P(X_0 = i) = 1 \quad \text{y} \quad P(Y_0 = j_0) = \pi_0(j_0), \quad j_0 \in T$$

Ya que $\{X_n\}_{n \in \mathbb{N}}$ y $\{Y_n\}_{n \in \mathbb{N}}$ son cadenas de Markov con probabilidades de transición p_{ij} , se verá que

$$P(X_n = j) = p_{ij}^{(n)} \quad \text{y} \quad P(Y_n = j) = \pi(j), \quad j \in T \quad (2.27)$$

Luego de (2.26) y (2.27) se cumple que

$$\lim_{n \rightarrow \infty} (p_{ij}^{(n)} - \pi(j)) = \lim_{n \rightarrow \infty} (P(X_n = j) - P(Y_n = j)) = 0$$

Por tanto

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \pi(j)$$

□

En este capítulo hemos visto el estudio de las cadenas de Markov desde el punto de vista probabilístico. Se analizó el comportamiento asintótico y el tratamiento que le da la teoría de probabilidad a dichas cadenas. A continuación, en el capítulo siguiente, veremos el tratamiento que se le da a las cadenas de Markov desde el enfoque estadístico. Nos centraremos especialmente en el proceso de inferencia estadística sobre las probabilidades de transición de estas cadenas. Todo ello lo realizaremos mediante el método de verosimilitud.

Capítulo 3

Inferencia Estadística

Se refiere al estudio de cómo sacar (inferir) conclusiones generales para una población a partir del estudio de una muestra, y que tengan cierto grado de fiabilidad o significación de los resultados obtenidos. Los problemas más importantes que engloba este proceso son: estimación puntual de parámetros, estimación de parámetros por intervalos y prueba de hipótesis sobre parámetros.

En este capítulo se revisan algunos conceptos básicos de la teoría de inferencia estadística que serán utilizados a lo largo de esta tesis. Aquí se definen conceptos generales como: la función de verosimilitud, la función de verosimilitud relativa, la aproximación continua a la función de verosimilitud y sus propiedades, intervalos de verosimilitud y verosimilitud-confianza.

3.1. Función de verosimilitud

Consideremos una muestra de variables aleatorias discretas X_1, X_2, \dots, X_n independientes con la misma función de distribución, donde $X = (X_1, \dots, X_n)'$ tiene función de probabilidad $P(X = x)$ que depende de un número finito de parámetros reales desconocidos $\theta = (\theta_1, \dots, \theta_m)' \in \Theta \subset \mathbb{R}^m$. El espacio parametral Θ es la región de valores posibles que puede tomar el vector de parámetros θ . Para resaltar que la función de probabilidad depende de parámetros desconocidos θ escribiremos $P(X = x; \theta)$.

Definición 3.1.1 *La función de verosimilitud del parámetro θ dado un conjunto de observaciones $x = (x_1, x_2, \dots, x_n)'$ se define como una función proporcional a la probabilidad de observar x cuando X sigue un modelo probabilístico paramétrico de θ*

$$L(\theta; x) \propto P(X = x; \theta) \tag{3.1}$$

donde x es un vector de observaciones de X y θ es el valor que indica cuán probable es observar x ([10, pág. 24]).

La función de verosimilitud juega un papel fundamental en la inferencia estadística. Su rol principal es inferir sobre los parámetros de la distribución que haya sido elegida para describir mejor al fenómeno aleatorio de interés a partir de una muestra observada. Notemos que esto es particularmente relevante después de un experimento, cuando ya fueron observadas las variables aleatorias. También es importante notar que la función de verosimilitud definida en (3.1) es proporcional y no igual, a la función de probabilidad $P(X = x; \theta)$. Esto enfatiza que solo el cociente de verosimilitudes

$$\frac{L(\theta_1; x)}{L(\theta_2; x)} = \frac{P(X = x; \theta_1)}{P(X = x; \theta_2)}$$

tiene significado, y se interpreta como una medida de la plausibilidad de θ_1 relativa a θ_2 basada en la muestra observada x . El cociente $L(\theta_1; x)/L(\theta_2; x) = k$ significa que el valor de θ_1 es k veces más plausible que el valor de θ_2 en el sentido de que θ_1 hace a la muestra observada k veces más probable de lo que la hace θ_2 . Es por ello que se recomienda usar a la función de verosimilitud relativa que definiremos a continuación para hacer inferencia sobre los parámetros del modelo probabilístico $P(X = x; \theta)$.

3.2. Función de verosimilitud relativa

Para que la función de verosimilitud tenga una única representación, que no involucre una constante arbitraria, es conveniente estandarizar a la función de verosimilitud con respecto a su máximo. El resultado de esto es la función de verosimilitud relativa ([33, pág. 9]).

Definición 3.2.1 *La función de verosimilitud relativa de θ se define como*

$$R(\theta; x) = \frac{L(\theta; x)}{\max_{\theta} L(\theta; x)} = \frac{L(\theta; x)}{L(\hat{\theta}; x)} \quad (3.2)$$

donde $\hat{\theta} = \hat{\theta}(x)$ es el valor del parámetro que maximiza $L(\theta; x)$ y se llama *estimador de máxima verosimilitud (EMV)* de θ .

El EMV $\hat{\theta}$ es el valor de θ más plausible, es decir, $\hat{\theta}$ es el valor de θ que hace más probable a la muestra observada. Puesto que $P(X = x; \theta)$ es una probabilidad, necesariamente está acotada entre 0 y 1 por lo que el denominador en (3.2) siempre existe y es finito. Así, la función de verosimilitud relativa yace entre cero y uno, $0 \leq R(\theta; x) \leq 1$, para todo valor de θ en el espacio parametral. La función de verosimilitud relativa proporciona la plausibilidad

de cualquier valor especificado de θ relativo al máximo verosímil $\hat{\theta}$, basada en la muestra observada x . Valores de θ con $R(\theta; x)$ cercanos a uno son muy creíbles o plausibles mientras que valores cercanos a cero son poco creíbles a la luz de la muestra observada.

3.3. Propiedades de la verosimilitud

3.3.1. No aditividad de las verosimilitudes

A diferencia de la probabilidad, la verosimilitud no es aditiva. Esta es la principal distinción entre verosimilitud y probabilidad. La verosimilitud es una función puntual cuyo dominio es el espacio parametral y su contradominio son los números reales. En cambio la probabilidad es una función de conjuntos, el dominio es una sigma-álgebra y el contradominio es el intervalo $[0, 1]$. La probabilidad de la unión de conjuntos disjuntos A y B está bien definida, $P(A \cup B) = P(A) + P(B)$. Pero la verosimilitud de la unión de dos valores θ_1 y θ_2 en el espacio parametral no está definida. La unión de θ_1 y θ_2 no es un real y por tanto no puede ser argumento de una función de verosimilitud.

3.3.2. Combinación de experimentos

Las funciones de verosimilitud combinan datos de experimentos diferentes de manera muy simple. Como la probabilidad conjunta de eventos independientes es el producto de sus probabilidades individuales, entonces la función de verosimilitud de θ , definida en (3.1), basada en diferentes conjuntos de datos que provienen de eventos (experimentos) independientes es el producto de las verosimilitudes individuales basadas en cada uno de estos conjuntos de datos. Así, el logaritmo de las verosimilitudes individuales basadas en cada conjunto de datos independientes se combinan a través de su suma. En particular esto significa que la forma apropiada de combinar información de experimentos diferentes que involucran un parámetro común de interés θ es a través de la suma de los respectivos logaritmos de las verosimilitudes individuales y esta función se maximiza para encontrar $\hat{\theta}$ el *EMV* común a todos los experimentos. Notemos que cada experimento podría involucrar a otros parámetros de estorbo particulares.

3.3.3. Invarianza funcional

La invarianza funcional es una característica muy conveniente de las verosimilitudes. Significa que cualquier declaración cuantitativa acerca de θ implica una declaración cuantitativa correspondiente acerca de cualquier función uno a uno de θ , $\delta = \delta(\theta)$, por directa sustitución algebraica $\theta = \theta(\delta)$. Por ejemplo, si $\theta > 0$ y $\delta = \log \theta$, entonces la verosimilitud del nuevo parámetro δ es

$$R^*(\delta; x) = R[\theta = \exp(\delta); x]$$

Como consecuencia se tiene que el *EMV* de δ es $\hat{\delta} = \log \hat{\theta}$. También $a \leq \theta \leq b$ si y solo si $\log a \leq \delta \leq \log b$. Estas dos declaraciones son equivalentes debido a que tienen la misma plausibilidad o incertidumbre. La invarianza funcional de la verosimilitud es una propiedad muy útil en la práctica. En muchos casos ocurre que el parámetro θ no es de interés principal sino que lo es otro parámetro que es función de θ . Por otro lado, con frecuencia un cambio en el parámetro puede simetrizar la forma de la función de verosimilitud. Es decir, $R^*(\delta; x)$ puede ser más simétrica, o tener una forma aproximadamente más normal, que $R(\theta; x)$. Así, inferencias sobre δ tendrán una estructura más simple, pero matemáticamente equivalente, que aquellas en términos de θ ya que se “acelera” la cercanía a las propiedades asintóticas del *EMV* $\hat{\delta}$ para la muestra pequeña en cuestión ([33, págs. 34-35]).

3.4. Aproximación continua de la función de verosimilitud y la verosimilitud exacta

La función de verosimilitud en (3.1) se definió en términos de variables aleatorias discretas. Sin embargo, esto no involucra una pérdida de generalidad ya que en realidad los datos observados, x , siempre son discretos puesto que todo instrumento de medición tiene precisión finita, y solo pueden registrarse mediciones con un número finito de decimales ([2]).

Cuando X_i es una variable aleatoria continua, la observación $X_i = x_i$ debe interpretarse como $X_i \in [x_i - \frac{h}{2}, x_i + \frac{h}{2}]$, donde h es un número positivo fijo que representa la precisión del instrumento de medición. Entonces, para una muestra de variables aleatorias independientes e idénticamente distribuidas $X = (X_1, \dots, X_n)'$ con función de densidad $f(x; \theta)$, la función de verosimilitud de θ es proporcional a la probabilidad conjunta de la muestra observada

$$\begin{aligned} L(\theta; x) &\propto \prod_{i=1}^n P\left(x_i - \frac{h}{2} \leq X_i \leq x_i + \frac{h}{2}; \theta\right) \\ &= \prod_{i=1}^n \int_{x_i - \frac{h}{2}}^{x_i + \frac{h}{2}} f(x_i; \theta) dx \end{aligned} \quad (3.3)$$

A la función de verosimilitud de θ definida en (3.3) la llamaremos **verosimilitud exacta** de θ . Notemos que $h = 0$ significa que el instrumento de medición tiene precisión infinita y que las observaciones pueden registrarse con un número infinito de decimales, lo cual es imposible en la realidad. Para una variable aleatoria continua X se tiene que $P(X = x; \theta) = 0$ para todo x y θ . Por ello, no se puede definir a la función de verosimilitud a partir de dichas probabilidades en el caso de variables continuas.

Si en contraste, se supone que la precisión del instrumento de medición es $h > 0$, entonces bajo ciertas condiciones la función de densidad $f(x; \theta)$ podrá usarse como una aproximación a la función de verosimilitud exacta (3.3). Sin embargo, si la función de densidad tiene una singularidad (discontinuidad infinita) en cualquier valor de θ , entonces $\max_{\theta} L(\theta; x)$, el denominador en (3.2), es infinito y entonces $R(\theta; x)$ en (3.2) no estaría definida. Este sería un caso donde la densidad $f(x; \theta)$ no podría usarse para aproximar a la función de verosimilitud.

Sin embargo, la aproximación de la verosimilitud exacta, a través de una densidad, es justificable bajo ciertas condiciones. Por el teorema del valor medio para integrales de funciones continuas, la i -ésima integral en (3.3) es $hf(x'; \theta)$ para algún punto intermedio $x' \in [x_i - \frac{h}{2}, x_i + \frac{h}{2}]$. Si $f(x; \theta)$ es aproximadamente constante en este intervalo para todo valor plausible de θ , entonces $f(x'; \theta) \approx f(x; \theta)$ en ese mismo intervalo. Si esta aproximación es adecuada para algunos o todos los $i \in \{1, \dots, n\}$ y si h no depende de θ entonces las correspondientes probabilidades en (3.3) pueden reemplazarse por la función de densidad evaluada en el valor observado x_i . En la práctica, es usual que esta aproximación se use y sustituya a todas las probabilidades en (3.3). Esto es

$$L(\theta; x) \propto \prod_{i=1}^n P\left(x_i - \frac{h}{2} \leq X_i \leq x_i + \frac{h}{2}\right) \approx \prod_{i=1}^n hf(x_i; \theta) \propto \prod_{i=1}^n f(x_i; \theta)$$

Kalbfleisch ([15, sección 9.4]), Edwards ([7, pág. 6, pág. 167]), Lindsey ([19]), Sprott ([33, pág. 19, págs. 203-294]), Lawless ([17, pág. 186]), Meeker y Escobar ([24, pág. 275]) discuten sobre lo necesario y lo importante que es tomar en cuenta estos argumentos. Hay que enfatizar que (3.3) no es el intento ad hoc para discretizar una variable aleatoria continua x , con función de densidad $f(x; \theta)$, la cual se considera como la base para calcular la función de verosimilitud. En realidad se trata de lo opuesto, una aproximación continua a una variable aleatoria discreta x , con función de probabilidad $P(X = x; \theta)$, la cual es la base para calcular la función de verosimilitud. El propósito principal de la aproximación continua es la conveniencia matemática; derivadas e integrales son más fáciles de calcular que diferencias y sumas finitas.

3.5. Intervalos de verosimilitud

Una manera usual de hacer inferencia sobre un parámetro de interés es a través de intervalos o regiones de estimación. Los intervalos de verosimilitud, o de forma más general de regiones de verosimilitud, indican los valores más plausibles del parámetro a la luz de la muestra observada. A continuación definiremos los intervalos de verosimilitud para un parámetro θ .

Definición 3.5.1 *Un intervalo de verosimilitud o región de verosimilitud de nivel c para θ , se define como*

$$IV(c) = \{\theta : R(\theta; x) \geq c\} \quad (3.4)$$

donde $0 \leq c \leq 1$.

Todo valor de θ en el $IV(c)$ tiene verosimilitud relativa igual o mayor que c , y todo valor de θ afuera, tiene verosimilitud relativa menor. Por tanto el $IV(c)$ separa los valores plausibles de θ de los no plausibles a un nivel c ([33, pág. 14]).

Cuando θ es unidimensional, el $IV(c)$ se obtiene trazando una línea horizontal en la gráfica de $R(\theta; x)$ a una distancia c paralela al eje cartesiano θ . Además, variando c de 0 a 1 se obtiene una familia jerarquizada y anidada de intervalos de verosimilitud que converge al $EMV \hat{\theta}$ cuando c tiende a 1. El $EMV \hat{\theta}$ está contenido en todos los intervalos de verosimilitud puesto que $R(\hat{\theta}; x) = 1$. Así, esta familia de intervalos anidados es equivalente a la función de verosimilitud completa y reproduce la gráfica de $R(\theta; x)$.

Un intervalo de verosimilitud por sí solo no es muy informativo y por lo tanto insuficiente para indicar el cambio en la plausibilidad de los valores de θ adentro del intervalo. Al menos, un intervalo de verosimilitud debe estar acompañado siempre del valor del $EMV \hat{\theta}$ para dar alguna idea de la simetría de la función de verosimilitud con respecto a $\hat{\theta}$ y de cómo cambia la plausibilidad adentro del intervalo. Por ello, se recomienda dar el EMV de θ junto con varios intervalos de verosimilitud de nivel $c = 0,036, 0,15$ y $0,25$. En la siguiente sección se verá que estos niveles de verosimilitud están asociados a confianzas aproximadas del 99 %, 95 % y 90 %, respectivamente.

3.6. Probabilidad de cobertura de intervalos de verosimilitud

Los resultados que veremos en esta sección son de gran relevancia al momento de asociar un nivel de confianza a los intervalos de verosimilitud. Lo haremos a través de la convergencia en distribución del estadístico de prueba basado en el método de máxima verosimilitud, como lo es el estadístico de razón de verosimilitud (Neyman y Pearson, 1928) $\Lambda = -2 \log R(\theta_0; x)$, a una variable aleatoria chi-cuadrado con un grado de libertad. Para la demostración de este resultado citaremos algunos antecedentes relevantes de los cuales no se incluye la demostración, para una mayor referencia de su desarrollo cito a Serfling ([31]).

Primeramente debemos considerar condiciones de regularidad para la familia de funciones de probabilidad que a continuación describiremos.

Sea Θ un intervalo abierto, no necesariamente finito ([31]), se cumplen las siguientes condiciones de regularidad:

Condiciones de regularidad

i. Para $\theta \in \Theta$, las derivadas

$$\frac{\partial \log f(x; \theta)}{\partial \theta}, \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} \text{ y } \frac{\partial^3 \log f(x; \theta)}{\partial \theta^3}$$

existen para todo x .

ii. Para $\theta_0 \in \Theta$ existen funciones $F(x)$, $G(x)$ y $H(x)$, tal que para θ en una vecindad de θ_0 se cumple que

$$\left| \frac{\partial \log f(x; \theta)}{\partial \theta} \right| \leq F(x), \left| \frac{\partial^2 \log f(x; \theta)}{\partial \theta^2} \right| \leq G(x) \text{ y } \left| \frac{\partial^3 \log f(x; \theta)}{\partial \theta^3} \right| \leq H(x)$$

para todo x . Además

$$\int F(x) dx < \infty, \int G(x) dx < \infty \text{ y } E_{\theta}[H(x)] < \infty$$

iii. Para cada $\theta \in \Theta$, se cumple que

$$0 < E_{\theta} \left[\frac{\partial \log f(x; \theta)}{\partial \theta} \right]^2 < \infty$$

La condición i asegura que la función $\frac{\partial \log f(x; \theta)}{\partial \theta}$ tiene, para cada x , una expansión de Taylor como una función de θ . La condición ii asegura que las funciones $\int f(x; \theta) dx$ y $\int \left[\frac{\partial \log f(x; \theta)}{\partial \theta} \right] dx$ pueden ser diferenciables con respecto a θ bajo la integral. La condición iii establece que las variables aleatorias $\frac{\partial \log f(x; \theta)}{\partial \theta}$ tienen varianza finita positiva.

Otro antecedente que nos será de utilidad en la demostración es la siguiente proposición:

Proposición 3.6.1 *Bajo las condiciones de regularidad i, ii y iii, consideremos observaciones independientes e idénticamente distribuidas en la familia de funciones de probabilidad, para $\theta \in \Theta$. Luego, con probabilidad 1, las ecuaciones de verosimilitud admiten una sucesión $\{\hat{\theta}_n\}$ de estimadores que satisfacen*

- a. *Consistencia, en el caso de poblaciones infinitas se dice que un estimador $\hat{\theta}_n$ es consistente si $(\hat{\theta}_n - \theta) \xrightarrow{P} 0$, cuando $n \rightarrow \infty$.*
- b. *Asintocidad normal, $\sqrt{n}(\hat{\theta}_n - \theta)$ converge en distribución a una normal con media 0 y varianza $I^{-1}(\theta; x)$, es decir, $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, I^{-1}(\theta; x))$.*

Por último, utilizaremos en la demostración el lema de Slutsky relacionado con convergencia en distribución de sucesiones de variables aleatorias.

Lema 3.6.1 Sean $\{X_n\}$ e $\{Y_n\}$ sucesiones de variables aleatorias tales que $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{P} b$, donde b es una constante. Entonces

$$\begin{aligned} X_n + Y_n &\xrightarrow{d} X + b \\ X_n Y_n &\xrightarrow{d} bX \\ \frac{X_n}{Y_n} &\xrightarrow{d} \frac{X}{b}, \text{ si } b \neq 0 \end{aligned}$$

Para demostrar la convergencia del estadístico de razón de verosimilitud lo haremos mediante una hipótesis nula simple del tipo $H_0 : \theta = \theta_0$. Con este resultado determinaremos para qué niveles de c , los intervalos de verosimilitud se relacionan con ciertos niveles de significancia, como por ejemplo $\alpha = 0,1, 0,05$ y $0,01$, a través de los cuantiles de la variable aleatoria chi-cuadrado con un grado de libertad, proporcionando de esta manera, una probabilidad de cobertura para los intervalos de verosimilitud para los mencionados niveles de α .

Teorema 3.6.1 Bajo las condiciones de regularidad *i, ii y iii*, y bajo la hipótesis $H_0 : \theta = \theta_0$, el estadístico de razón de verosimilitud $\Lambda = -2 \log R(\theta_0; x)$ converge en distribución a una chi-cuadrado con un grado de libertad, $\chi^2_{(1)}$.

Demostración

Sea $l_n(\theta) = \log L(\theta; x)$ la función de logverosimilitud. Construyamos ahora la expansión de Taylor de $l_n(\theta)$ alrededor de $\theta = \hat{\theta}_n$

$$l_n(\theta_0) = l_n(\hat{\theta}_n) + (\hat{\theta}_n - \theta_0)l'_n(\hat{\theta}_n) + \frac{(\hat{\theta}_n - \theta_0)^2}{2!}l''_n(\hat{\theta}_n) + R_1(\theta_0, \hat{\theta}_n)$$

Por ser $\hat{\theta}$ el *EMV*, entonces $l'_n(\hat{\theta}_n) = 0$. Reemplazando este resultado en la ecuación anterior tendríamos

$$l_n(\theta_0) - l_n(\hat{\theta}_n) = \frac{(\hat{\theta}_n - \theta_0)^2}{2!}l''_n(\hat{\theta}_n) + R_1(\theta_0, \hat{\theta}_n)$$

Luego

$$l_n(\theta_0) - l_n(\hat{\theta}_n) = n(\hat{\theta}_n - \theta_0)^2 I(\theta_0) \left[\frac{l''_n(\hat{\theta}_n)}{2nI(\theta_0)} + R_2(\theta_0, \hat{\theta}_n) \right]$$

de donde obtenemos

$$\Lambda = -2 \log R(\theta_0; x) = -2[l_n(\theta_0) - l_n(\hat{\theta}_n)] = n(\hat{\theta}_n - \theta_0)^2 I(\theta_0) \left[\frac{l_n''(\hat{\theta}_n)}{nI(\theta_0)} + R_3(\theta_0, \hat{\theta}_n) \right]$$

Considerando que bajo $H_0 : \theta = \theta_0$, se cumple que

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \sqrt{I(\theta_0)} \xrightarrow{d} N(0, 1) \implies n(\hat{\theta}_n - \theta_0)^2 I(\theta_0) \xrightarrow{d} \chi_{(1)}^2$$

Por la ley fuerte de los grandes números tenemos

$$\frac{l_n''(\theta_0)}{n} \xrightarrow{c.s.} I(\theta_0) \implies \frac{l_n''(\hat{\theta}_n)}{nI(\theta_0)} \xrightarrow{P} 1$$

Por otra parte, por la consistencia $(\hat{\theta}_n - \theta) \xrightarrow{P} 0$, cuando $n \rightarrow \infty$, tenemos que

$$R_3(\theta_0, \hat{\theta}_n) \xrightarrow{P} 0$$

Luego, por el lema de Slutsky

$$\Lambda \xrightarrow{d} \chi_{(1)}^2$$

Por lo tanto, el estadístico de razón de verosimilitud converge en distribución a una variable aleatoria chi-cuadrado con un grado de libertad.

□

Ahora, supongamos que tenemos una muestra de variables aleatorias independientes e idénticamente distribuidas $X = (X_1, \dots, X_n)'$ con función de probabilidad $P(X = x; \theta)$, donde θ es un parámetro escalar desconocido. Si se tiene una muestra observada y que proviene de la distribución de X con θ fijo en un valor θ_0 , entonces se puede calcular, a partir de esta muestra observada, un intervalo $[A, B]$ para el valor verdadero θ_0 .

Definición 3.6.1 *La probabilidad de cobertura de un intervalo aleatorio $[A, B]$ es la probabilidad de que el intervalo $[A, B]$ incluya o cubra, el verdadero valor del parámetro θ_0*

$$PC(\theta_0) = P(A \leq \theta_0 \leq B; \theta = \theta_0) \quad (3.5)$$

La probabilidad de cobertura $PC(\theta_0)$ se interpreta como la fracción de veces que el intervalo $[A, B]$ incluirá el valor verdadero θ_0 en un número muy grande de repeticiones de la muestra (o experimento) pero con el valor de θ fijo en θ_0 . Aunque existen definiciones más

generales de probabilidad de cobertura, adoptaremos esta definición porque es suficiente y clara para los objetivos de este trabajo.

Los extremos del intervalo $[A, B]$ son variables aleatorias puesto que su valor varía cuando cambia la muestra. Observemos que en principio, la distribución de probabilidad de A y B se puede calcular a partir de la distribución de la variable aleatoria X y generalmente depende de θ_0 . Además, como el intervalo $[A, B]$ puede variar cuando varía la muestra, entonces este puede algunas veces incluir y en otras no, al valor verdadero θ_0 .

Definición 3.6.2 *El intervalo aleatorio $[A, B]$ es un intervalo de confianza (IC) para θ si la probabilidad de cobertura $PC(\theta_0) = P(A \leq \theta_0 \leq B; \theta = \theta_0)$ es la misma para todos los valores θ_0 del parámetro, esto es, cuando dicha probabilidad de cobertura no depende del parámetro θ_0 ([15. pág. 200]).*

En primera instancia, $[A, B]$ es un intervalo del 95 % para θ si $PC(\theta_0) = P(A \leq \theta_0 \leq B; \theta = \theta_0) = 0,95$ para todos los posibles valores de θ_0 . Un IC del 95 % podría incluir el valor verdadero θ_0 del parámetro en un 95 % de las repeticiones del experimento.

La probabilidad de cobertura de un $IV(c)$ se puede calcular a través de la distribución de probabilidad de la estadística de la razón de verosimilitud Λ para un θ fijo en θ_0 , $\Lambda \equiv -2 \log R(\theta_0; x)$. Un valor particular θ_0 está en el $IV(c)$ si y solo si $R(\theta_0; x) \geq c$, o de forma equivalente, $-2 \log R(\theta_0; x) \leq -2 \log c$. Por lo tanto, la probabilidad de cobertura del $IV(c)$ es

$$PC(\theta_0) = P[\theta_0 \in IV(c); \theta = \theta_0]$$

$$PC(\theta_0) = P(\Lambda \leq -2 \log c; \theta = \theta_0) \tag{3.6}$$

En general es difícil encontrar la distribución de probabilidad exacta de Λ ; sin embargo por el teorema 3.6.1 se cumple que el estadístico de la razón de verosimilitud $\Lambda = -2 \log R(\theta_0; x)$ converge a una chi-cuadrado con un grado de libertad, $\chi^2_{(1)}$, para todo $\theta \in \Theta \subset \mathbb{R}$. Luego de (3.6) se tiene que $PC(\theta_0) = P(\Lambda \leq d; \theta = \theta_0)$ es la probabilidad de cobertura del $IV(c)$.

Por (3.6), $P(\Lambda \leq d)$ es la cobertura del $IV(c)$, donde $d = -2 \log c$; ya que $P(\Lambda \leq d = -2 \log c)$, entonces $c = e^{-\frac{d}{2}}$. Si se selecciona un valor de tablas $d = q(\alpha, 1) = 3,841$, donde $q(\alpha, 1)$ es el cuantil $(1 - \alpha)$ de una chi-cuadrado con un grado de libertad, tenemos que $c = e^{-\frac{3,841}{2}} = 0,147$, por tanto $P(\Lambda \leq 3,841)$ es la probabilidad de cobertura del $IV(c)$ con $c = 0,147$.

Observamos de la tabla 3.1. que los valores 2,706, 3,841 y 6,635 son los cuantiles 0,90, 0,95 y 0,99 de una distribución chi-cuadrado con un grado de libertad, respectivamente, por lo que los $IV(c)$ con $c = 0,258$, 0,147 y 0,036 tiene una probabilidad de cobertura aproximada

$1 - \alpha$	c	$d = q(\alpha, 1)$
0,90	0,258	2,706
0,95	0,147	3,841
0,99	0,036	6,635

Tabla 3.1: Confianza aproximada de $IV(c)$ cuando θ es unidimensional

de 90 %, 95 % y 99 %, respectivamente. De esta manera, $IV(c)$ también es un intervalo o región de confianza para el parámetro θ y toma, en consecuencia, el nombre de *intervalo de verosimilitud-confianza* para θ .

La mejor manera de construir un intervalo de confianza es a partir de la función de verosimilitud, con la finalidad de no caer en el caso que se estén rechazando valores plausibles por no ser cubiertos por el intervalo. Cuando aplicamos un estadístico de razón de verosimilitud $\Lambda = -2 \log R(\theta_0; x) \approx \chi_{(1)}^2$, el coeficiente de confianza aproximado, es decir, la probabilidad de cobertura del $IV(c)$ está dada por

$$PC \approx P(\chi_{(1)}^2 \leq -2 \log c)$$

porque de la aproximación chi-cuadrado, los intervalos de verosimilitud son exactos o se aproximan a los intervalos de confianza en la mayoría de las aplicaciones. Como el cuadrado de una variable chi-cuadrado con un grado de libertad es una normal estándar, por el inciso b de la proposición 3.6.1, $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0; x))$, tenemos

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta_0)}{\sqrt{I^{-1}(\theta_0; x)}} \xrightarrow{d} N(0, 1)$$

Por el lema 3.6.1 de Slutsky

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta_0)}{\sqrt{I^{-1}(\hat{\theta}_n; x)}} \xrightarrow{d} N(0, 1)$$

donde $\sqrt{I^{-1}(\hat{\theta}_n; x)}$ es un pivotal ya que es una variable aleatoria que no depende de θ . Esto es, es constante para todo θ .

Ejemplo 3.6.1 Consideremos un cierto tipo de componente electrónico que falla en cualquier instante de tiempo. Sin embargo, los componentes no se deterioran con la edad, y la probabilidad de falla dentro de un periodo de tiempo dado no depende de la edad del componente. Se supone que el tiempo de vida de tales componentes sigue una distribución exponencial $F(x; \theta)$ con función de densidad

$$f(x; \theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \quad \text{para todo } x \text{ positivo}$$

donde θ es el tiempo de vida esperado de los componentes. El objetivo es hacer inferencia sobre el parámetro θ basados en observaciones de tiempo de vida de los componentes.

Supongamos que n componentes electrónicos son puestos a prueba durante periodos de tiempo fijo T_1, \dots, T_n . Consideremos que r de estos componentes fallaron antes de culminar sus periodos de prueba, y que los tiempos de falla observados fueron t_1, \dots, t_r . Así, los $(n - r)$ componentes restantes no fallaron durante sus periodos de prueba y tuvieron tiempos de vida censurados T_{r+1}, \dots, T_n . Entonces, usando la aproximación continua de la función de verosimilitud, solo en los r componentes que fallaron, y considerando el aporte de los $(n - r)$ tiempos de vida censurados se tiene que la función de verosimilitud de θ es

$$\begin{aligned} L(\theta; x) &\propto \left[\prod_{i=1}^r f(t_i; \theta) \right] \left[\prod_{i=r+1}^n (1 - F(T_i; \theta)) \right] \\ &= \left[\prod_{i=1}^r \frac{1}{\theta} \exp\left(-\frac{t_i}{\theta}\right) \right] \left[\prod_{i=r+1}^n \exp\left(-\frac{T_i}{\theta}\right) \right] \\ &= \theta^{-r} \exp\left(-\frac{t}{\theta}\right), \quad \text{donde } t = \sum_{i=1}^r t_i + \sum_{i=r+1}^n T_i \end{aligned}$$

Para encontrar el EMV de θ procederemos de la siguiente manera. Tomando el logaritmo natural obtenemos la función de logverosimilitud de θ

$$\log L(\theta; x) = -r \log \theta - \frac{t}{\theta}$$

Derivando e igualando a cero, obtenemos un punto crítico

$$\theta = \frac{t}{r}$$

Utilizando el criterio de la segunda derivada verificaremos si este punto crítico maximiza a la función de verosimilitud de θ . Para $x > 0$ tenemos que

$$\begin{aligned} \frac{\partial^2 \log L(\theta; x)}{\partial \theta^2} &= \frac{r}{\theta^2} - \frac{2t}{\theta^3} \\ \frac{\partial^2 \log L(\theta; x)}{\partial \theta^2} \Big|_{\theta=t/r} &= \frac{r^2}{t^2} \left(r - \frac{2t}{t/r} \right) < 0 \end{aligned}$$

Entonces la función de verosimilitud de θ tiene un máximo local en $\theta = \frac{t}{r}$. Por tanto, el EMV de θ es $\hat{\theta} = \frac{t}{r}$.

Utilizando la definición de verosimilitud relativa dada en (3.2), tenemos

$$R(\theta; x) = \frac{L(\theta; x)}{L(\hat{\theta}; x)}$$

$$R(\theta; x) = \left(\frac{\theta}{\hat{\theta}}\right)^{-r} \exp\left(-\frac{t}{\theta} + \frac{t}{\hat{\theta}}\right)$$

Sustituyendo el valor del EMV en la expresión anterior, la función de verosimilitud queda expresada como

$$R(\theta; x) = \left(\frac{r}{t}\theta\right)^{-r} \exp\left(-\frac{t}{\theta} + r\right)$$

Un ejemplo que aparece en la literatura es $n = 10$, con periodos de tiempo de prueba fijo $T = 81, 70, 41, 31, 31, 30, 29, 72, 60, 21$ días. De estos componentes, siete fallaron antes de culminar su periodo de prueba, y los tiempos de falla observados fueron $t = 2, 51, 33, 27, 14, 24, 4$ días. Los tres tiempos de vida censurados fueron $T_8 = 72$, $T_9 = 60$ y $T_{10} = 21$ (Bartholomew, 1957 y [33, pág. 23]). Así, como $r = 7$ y $t = 308$ entonces la función de verosimilitud relativa de θ es

$$R(\theta; x) = \left(\frac{7}{308}\theta\right)^{-7} \exp\left(-\frac{308}{\theta} + 7\right) \quad (3.7)$$

La figura 3.1 muestra la gráfica de $R(\theta; x)$ en (3.7), los intervalos de verosimilitud de nivel $c = 0, 036, 0, 15$ y $0, 25$ y la ubicación del EMV de θ , $\hat{\theta} = 252/8 = 31,5$. Se observa claramente que la función de verosimilitud relativa de θ es asimétrica con cola pesada a la derecha. Esto también puede verse a partir de los intervalos de verosimilitud y de la ubicación del EMV $\hat{\theta}$. Existe una desviación evidente entre $\hat{\theta}$ y el centro geométrico de cada intervalo, lo que da idea de la asimetría de la función de verosimilitud.

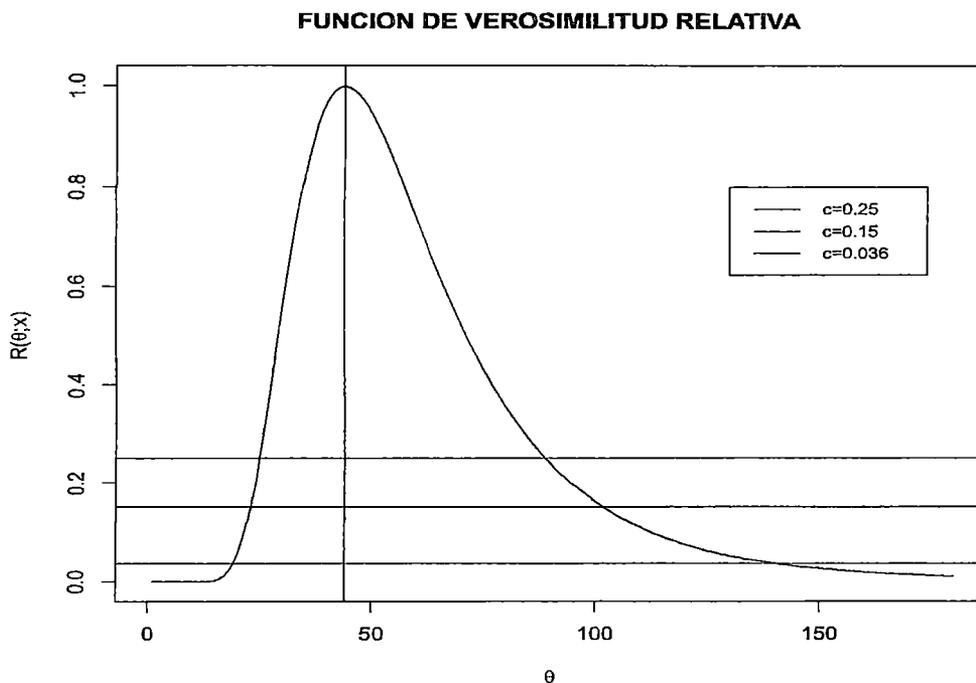


Figura 3.1: Intervalos de verosimilitud

El resumen de las inferencias en términos de verosimilitud consiste en proporcionar el nivel de verosimilitud, el extremo inferior del intervalo, A , el EMV $\hat{\theta}$ y el extremo superior B del intervalo de verosimilitud. Usualmente se dan intervalos de verosimilitud de niveles $c = 0,036, 0,15$ y $0,25$ porque suelen estar asociados con niveles del 99 %, 95 % y 90 % de confianza, en el caso de que el parámetro sea unidimensional.

c	A	$\hat{\theta}$	B
0,036	19,02	44	140,64
0,15	22,83	44	101,80
0,25	24,89	44	88,86

Tabla 3.2: Resumen de inferencias para θ

Notemos que valores de $\theta \leq 19,02$ y valores de $\theta \geq 140,64$ tienen plausibilidad muy pequeña, menor que $c = 0,036$.

Antes de finalizar esta sección es importante enfatizar que la verosimilitud es una función puntual y por lo tanto el nivel de plausibilidad c de un intervalo de verosimilitud no es una declaración de la incertidumbre del intervalo. Es una declaración acerca de la plausibilidad relativa de cualquier punto individual adentro del intervalo, cuya verosimilitud es mayor o igual a c .

Capítulo 4

Verosimilitud para Cadenas de Markov

Hemos dado los conceptos básicos de la teoría de probabilidad para cadenas de Markov y también la teoría de verosimilitud para inferencia estadística. Ahora veremos el enfoque de la verosimilitud aplicado a las cadenas de Markov, particularmente mostraremos el proceso inferencial para cadenas de Markov paramétricas y no paramétricas.

Consideremos una cadena de Markov $\{X_n\}_{n \in \mathbb{N}}$ irreducible con espacio de estados $T = \{1, 2, \dots, N\}$, probabilidades de transición p_{ij} , distribución inicial $\pi_0(i)$ y distribución estacionaria (límite) $\pi(j)$. Los parámetros a estimar son dichas probabilidades de transición y considerando los casos paramétrico y no paramétrico.

4.1. Caso paramétrico

En este caso las probabilidades de transición pueden ser indexadas por medio de un vector de parámetros. Supongamos que $p_{ij} = p_{ij}(\theta)$, donde $\theta = (\theta_1, \theta_2, \dots, \theta_r)'$ es un vector de parámetros desconocidos que toma valores en \mathbb{R}^r .

Sea $x = \{x_1, \dots, x_n\}$ una realización de observaciones de una cadena de Markov de tamaño n . Utilizando la probabilidad conjunta dada en (1.3), la función de verosimilitud para el caso paramétrico viene dada por

$$L(p_{ij}(\theta); x) \propto \pi_0(i_0) \prod_{k=1}^n p_{x_{k+1}x_k} = \pi_0(i_0) \prod_{i=1}^N \prod_{j=1}^N [p_{ij}(\theta)]^{n_{ij}} \quad (4.1)$$

donde n_{ij} son las frecuencias de las transiciones en un paso de ir de un estado i a un estado j . Tomando logaritmo natural tenemos la función de logverosimilitud

$$\log L(p_{ij}(\theta); x) = \log \pi_0(i_0) + \sum_{i=1}^N \sum_{j=1}^N n_{ij} \log p_{ij}(\theta) \quad (4.2)$$

Aquí podemos omitir el término $\log \pi_0(i_0)$ bajo las condiciones de regularidad **i**, **ii** y **iii** dadas en el capítulo 3 (Basawa, 1980, pág. 228).

Los estimadores de máxima verosimilitud de los parámetros $\theta_1, \theta_2, \dots, \theta_r$ se obtienen resolviendo el sistema de ecuaciones

$$\frac{\partial \log L(p_{ij}(\theta); x)}{\partial \theta_s} = 0, \quad s = 1, 2, \dots, r$$

Es decir

$$\sum_{i=1}^N \sum_{j=1}^N \left[n_{ij} \frac{\partial p_{ij}(\theta)}{\partial \theta_s} \cdot \frac{1}{p_{ij}(\theta)} \right] = 0, \quad s = 1, 2, \dots, r \quad (4.3)$$

para $\theta_1, \theta_2, \dots, \theta_r$.

Ejemplo 4.1.1 Con el fin de ejemplificar el caso anterior, supongamos que existe una cadena de Markov con espacio de estados $T = \{1, 2\}$ y matriz de probabilidades de transición

$$P = (p_{ij}(\theta)) = \begin{bmatrix} \theta & 1 - \theta \\ 1 - \theta^6 & \theta^6 \end{bmatrix} \quad (4.4)$$

Usaremos esta dependencia paramétrica (que puede resultar forzada), solo para fines prácticos.

De la ecuación (4.1), la función de verosimilitud para θ será

$$L(\theta; x) \propto [p_{11}(\theta)]^{n_{11}} [p_{22}(\theta)]^{n_{22}} [p_{12}(\theta)]^{n_{12}} [p_{21}(\theta)]^{n_{21}}$$

$$L(\theta; x) \propto \theta^{(n_{11} + 6n_{22})} (1 - \theta)^{n_{12}} (1 - \theta^6)^{n_{21}}$$

Ahora, usando la definición de verosimilitud relativa (3.2) tenemos

$$R(\theta; x) = \left(\frac{\theta}{\hat{\theta}} \right)^{(n_{11} + 6n_{22})} \left(\frac{1 - \theta}{1 - \hat{\theta}} \right)^{n_{12}} \left(\frac{1 - \theta^6}{1 - \hat{\theta}^6} \right)^{n_{21}}$$

A partir de datos reales de precipitación diaria discretizados en dos estados (seco=1, lluvioso=2) que fueron tomados de la estación meteorológica de Cerro de Pasco - 000593 (en la sección 6.2 se darán más detalles al respecto), se obtuvieron las frecuencias de las transiciones diarias totales para 5 años de precipitación (2009-2013): $n_{11} = 1136$, $n_{12} = 254$, $n_{21} = 254$ y $n_{22} = 180$.

A partir de la ecuación (4.3), podemos encontrar el EMV para el caso de un solo parámetro

$$\sum_{i=1}^2 \sum_{j=1}^2 \left[n_{ij} \frac{\partial p_{ij}(\theta)}{\partial \theta} \cdot \frac{1}{p_{ij}(\theta)} \right] = 0$$

Desarrollando la ecuación anterior tenemos

$$n_{11}(1) \cdot \frac{1}{\theta} + n_{12}(-1) \cdot \frac{1}{1-\theta} + n_{21}(-6\theta^5) \cdot \frac{1}{1-\theta^6} + n_{22}(6\theta^5) \cdot \frac{1}{\theta^6} = 0$$

Llevándolo a la forma polinomial

$$(n_{11} + n_{12} + 6n_{21} + 6n_{22})\theta^7 - (n_{11} + 6n_{21} + 6n_{22})\theta^6 - (n_{11} + n_{12} + 6n_{22})\theta + (n_{11} + 6n_{22}) = 0$$

Ahora, utilizando métodos numéricos, tenemos que el valor del EMV sería $\hat{\theta} = 0,843$.

También se determinaron los intervalos de verosimilitud-confianza a los niveles $c = 0,036$, $c = 0,15$ y $c = 0,25$. En la tabla 4.1 se muestran los resultados de inferencias de verosimilitud para θ :

c	A	$\hat{\theta}$	B
0,036	0,826	0,843	0,859
0,15	0,830	0,843	0,855
0,25	0,832	0,843	0,854

Tabla 4.1: Resumen de inferencias para el caso paramétrico

Si sustituimos el valor de $\hat{\theta}$ en (4.4), obtenemos la estimación puntual de la matriz P

$$\hat{P} = (p_{ij}(\hat{\theta})) = \begin{bmatrix} 0,843 & 0,157 \\ 0,641 & 0,359 \end{bmatrix} \quad (4.5)$$

La interpretación que le podemos dar a la matriz anterior es que si el día de hoy no llovió, entonces existe mayor probabilidad que mañana no llueva (0,843) de que llueva (0,157). Si por el contrario, el día de hoy llovió, entonces existe mayor probabilidad que mañana no llueva (0,641) de que llueva (0,359).

En la siguiente sección vamos a aprender a simular una cadena de Markov paramétrica y no paramétrica para poder verificar las probabilidades de cobertura asociadas a los intervalos de verosimilitud mediante este proceso.

4.1.1. Simulación de una cadena de Markov paramétrica

Mediante un estudio de simulación vamos a verificar las probabilidades de cobertura asociadas a los intervalos de verosimilitud. En esta parte haremos simulaciones de una cadena de Markov paramétrica y en la siguiente sección veremos el caso no paramétrico. A continuación mostramos el algoritmo para realizar esta simulación.

Algoritmo: Simulación de una cadena de Markov

Entrada

- π : distribución inicial
- P : matriz de probabilidades de transición de un paso
- n : número de datos a simular
- X : sucesión de ceros

Salida

- X

rando \leftarrow funcion(p)

$u \leftarrow \text{Uniforme}(0.1)$

$s \leftarrow p[1]$

$j \leftarrow 1$

Mientras ($u > s$) o ($j < \text{longitud de } p$)

$j \leftarrow j + 1$

$s \leftarrow s + p[j]$

Fin mientras

index $\leftarrow j$

Fin proceso

$X[1] \leftarrow \text{rando}(\pi)$

Para $i \leftarrow 1$ hasta n hacer

$$X[i + 1] \leftarrow \text{rando}(P[X[i], :])$$

Fin para

Escribir X

Fin proceso

Con el algoritmo anterior podemos obtener cadenas de Markov para el caso paramétrico y no paramétrico con las características que se deseen.

Realizaremos ahora la simulación considerando el ejemplo 4.1.1 para la cadena de Markov paramétrica con espacio de estados $T = \{1, 2\}$ y matriz de probabilidades de transición estimada dada en (4.5). Consideraremos una distribución estacionaria π_0 , para ello debemos resolver el sistema de ecuaciones dado por

$$\pi_0 \widehat{P} = \pi_0$$

que para este caso sería equivalente a resolver

$$\pi_0(1)\widehat{p}_{11} + \pi_0(2)\widehat{p}_{21} = \pi_0(1)$$

$$\pi_0(1)\widehat{p}_{12} + \pi_0(2)\widehat{p}_{22} = \pi_0(2)$$

con la condición

$$\pi_0(1) + \pi_0(2) = 1$$

Así tendríamos la distribución estacionaria dada por

$$\pi_0(1) = P(X_0 = 1) = 0,803 \text{ y } \pi_0(2) = P(X_0 = 2) = 0,197$$

Simularemos una cadena de Markov paramétrica con las características mencionadas y tomaremos una muestra de esta de tamaño $n = 1826$ para realizar la inferencia a través de la verosimilitud.

Puesto que en la ecuación (4.1) aparece el término n_{ij} (que son las frecuencias de las transiciones en un paso para ir del estado i al j en toda la muestra), colocaremos estos valores en una matriz F llamada *matriz de frecuencias*.

Una vez que tenemos la matriz de frecuencias F realizamos la inferencia basada en verosimilitud al igual que en el ejemplo 4.1.1

Los resultados del valor estimado $\hat{\theta}$ y de los intervalos de verosimilitud los mostramos a continuación en la siguiente tabla:

c	A	$\hat{\theta}$	B
0,036	0,830	0,848	0,865
0,15	0,834	0,848	0,861
0,25	0,836	0,848	0,859

Tabla 4.2: Resumen de inferencias para la simulación paramétrica

A partir de la tabla 4.2 podemos determinar las probabilidades de transición estimadas reemplazando el valor de $\hat{\theta}$ en la matriz del ejemplo 4.1.1.

$$\hat{P} = (p_{ij}(\hat{\theta})) = \begin{bmatrix} 0,848 & 0,152 \\ 0,628 & 0,372 \end{bmatrix} \quad (4.6)$$

Al igual que el ejemplo 4.1.1, la interpretación que le podemos dar a la matriz anterior es que si el día de hoy no llovió, entonces existe mayor probabilidad que mañana no llueva (0,848) de que llueva (0,152). Si por el contrario, el día de hoy llovió, entonces existe mayor probabilidad que mañana no llueva (0,628) de que llueva (0,372).

Como queremos verificar la confianza asociada a los intervalos de verosimilitud, veremos si el intervalo cubre o no el verdadero valor del parámetro ($\hat{\theta} = 0,843$) a los niveles de plausibilidad $c = 0,036, 0,15$ y $0,25$. Para ello realizaremos el proceso anterior 5000 veces y estimaremos las probabilidades de cobertura.

c	$1 - \alpha$	$PC(\theta)$
0,036	0,99	0,991
0,15	0,95	0,950
0,25	0,90	0,903

Tabla 4.3: Probabilidades de cobertura a los diferentes niveles de plausibilidad

De la tabla 4.3 vemos que la probabilidades de cobertura (nivel de confianza real) a los diferentes niveles de plausibilidad ($c = 0,036, 0,15$ y $0,25$) son muy similares a lo teóricamente indicado (nivel de confianza nominal). Se puede optar por usar otra distribución inicial π_0 , no necesariamente estacionaria, veremos que los resultados de las probabilidades de cobertura

son muy similares.

Pasaremos ahora a ver el caso paramétrico, pues el caso anterior resulta a veces complicado de modelar para un caso real.

4.2. Caso no paramétrico

Para este caso las probabilidades de transición p_{ij} no pueden ser indexadas a través de un vector de parámetros. Sea $x = \{x_1, x_2, \dots, x_n\}$ una realización de observaciones de una cadena de Markov de tamaño n . Entonces, utilizando la probabilidad conjunta dada en (2.1), la función de verosimilitud para el caso no paramétrico viene dada por

$$L(p_{ij}; x) \propto P(x; p_{ij}) = \pi_0(i_0) \prod_{k=1}^n p_{x_{k+1}x_k} = \pi_0(i_0) \prod_{i=1}^N \prod_{j=1}^N p_{ij}^{n_{ij}} \quad (4.7)$$

donde n_{ij} son las frecuencias de las transiciones en un paso de ir de un estado i a un estado j . Dichos números de transiciones serán colocados en una matriz $F = (n_{ij})$ llamada *matriz de frecuencias*. Debemos maximizar la función de verosimilitud para encontrar el *EMV* de p_{ij} . Para ello, tomando el logaritmo a la función de verosimilitud, obtenemos

$$\log L(p_{ij}; x) = \log \pi_0(i_0) + \sum_{i=1}^N \sum_{j=1}^N n_{ij} \log p_{ij} \quad (4.8)$$

sujeto a la restricción $\sum_{j=1}^N p_{ij} = 1$ y utilizando multiplicadores de Lagrange, obtenemos la siguiente expresión:

$$l^*(p_{ij}; x) = \log \pi_0(i_0) + \sum_{i=1}^N \sum_{j=1}^N n_{ij} \log p_{ij} + \lambda \left(1 - \sum_{j=1}^N p_{ij} \right) \quad (4.9)$$

Derivando l^* respecto a las p_{ij} y a λ , tenemos que

$$\frac{\partial l^*}{\partial p_{ij}} = \frac{n_{ij}}{p_{ij}} - \lambda \quad \text{y} \quad \frac{\partial l^*}{\partial \lambda} = 1 - \sum_{j=1}^N p_{ij}$$

Igualando las expresiones anteriores a cero, obtenemos

$$\lambda = \frac{n_{ij}}{p_{ij}} \quad \implies \quad p_{ij} = \frac{n_{ij}}{\lambda}$$

Sumando y usando que $\sum_{j=1}^N p_{ij} = 1$, deducimos que

$$\sum_{j=1}^N p_{ij} = \frac{\sum_{j=1}^N n_{ij}}{\lambda} \implies 1 = \frac{\sum_{j=1}^N n_{ij}}{\lambda} \implies \sum_{j=1}^N n_{ij} = \lambda$$

Tomando $\lambda = n_i$, obtenemos el *EMV*, a saber

$$\widehat{p}_{ij} = \frac{n_{ij}}{n_i} \quad (4.10)$$

donde $n_i = \sum_{j=1}^N n_{ij}$.

Nuevamente omitiremos el término $\log \pi_0(i_0)$ bajo las condiciones de regularidad **i**, **ii** y **iii** dadas en el capítulo 3 (Basawa, 1980, pág. 228).

Para hacer inferencias sobre las probabilidades de transición p_{ij} consideraremos la distribución marginal de n_{ij} . Sea la matriz de frecuencias

$$F = \begin{bmatrix} n_{11} & n_{12} & \cdots & n_{1j} & \cdots & n_{1N} \\ n_{21} & n_{22} & \cdots & n_{2j} & \cdots & n_{2N} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ n_{i1} & n_{i2} & \cdots & n_{ij} & \cdots & n_{iN} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ n_{N1} & n_{N2} & \cdots & n_{Nj} & \cdots & n_{NN} \end{bmatrix}$$

Si fijamos i en la matriz F , podemos suponer que

$$(n_{i1}, n_{i2}, \dots, n_{iN}) \sim \text{Multinomial}(n_i, p_{i1}, p_{i2}, \dots, p_{iN})$$

Y si ahora fijamos j tenemos que

$$n_{ij} \sim \text{Binomial}(n_i, p_{ij})$$

Luego, la función de verosimilitud de p_{ij} basada en la distribución marginal de n_{ij} es

$$L(p_{ij}) \propto p_{ij}^{n_{ij}} (1 - p_{ij})^{n_i - n_{ij}} \quad (4.11)$$

donde p_{ij} es una entrada de la matriz de probabilidades de transición P , n_{ij} es una entrada de la matriz de frecuencias F y n_i es la suma de las frecuencias del i -ésimo renglón. Con la función de distribución de p_{ij} basada en la distribución marginal de n_{ij} podemos encontrar la estimación de un parámetro específico cuando se tienen varios. Como deseamos hallar los intervalos de verosimilitud para las p_{ij} , necesitamos para ello la función de verosimilitud relativa

$$R(p_{ij}; x) = \frac{L(p_{ij}; x)}{L(\widehat{p}_{ij}; x)} = \left(\frac{p_{ij}}{\widehat{p}_{ij}} \right)^{n_{ij}} \left(\frac{1 - p_{ij}}{1 - \widehat{p}_{ij}} \right)^{n_i - n_{ij}} \quad (4.12)$$

Con la ecuación (4.12), podemos encontrar los intervalos de verosimilitud para cada una de las probabilidades de transición p_{ij} a diferentes niveles de c .

Pasaremos ahora a mostrar un ejemplo de aplicación real del caso no paramétrico.

4.3. Ejemplo de aplicación real

En esta sección pondremos en práctica el método propuesto mediante un ejemplo de aplicación real; para ello trabajaremos con datos reales de precipitación diaria del distrito de Chaupimarca ubicado en la provincia de Pasco, en el departamento del mismo nombre. Para definir el tipo y diseño de investigación nos basaremos en la siguiente cita textual: “Los datos longitudinales son muy comunes en investigación y usualmente surgen de medir una o varias respuestas, junto con características propias de cada sujeto o unidad experimental repetidamente a través del tiempo. Algunas de estas respuestas pueden pensarse como realizaciones aleatorias de un cierto proceso estocástico con espacio de estados discreto” (Tovar y Salazar, 2009, pág. 214) se trata pues de un estudio de tipo no experimental de diseño longitudinal, más precisamente de diseño longitudinal de evolución de grupo (cohorte) pues se trata de una muestra de estos datos y no de toda la población. A partir de los datos de precipitación diaria se ‘construirá’ una cadena de Markov de dos estados siendo cada uno de estos una condición de ocurrencia o no ocurrencia de precipitación (o lo que es lo mismo, si el día fue lluvioso o seco).

Los datos fueron tomados de la estación meteorológica Cerro de Pasco - 000593 ubicada en el distrito de Chaupimarca a 4260 msnm con coordenadas $10^{\circ} 41' 40''$ de latitud y $76^{\circ} 15' 15''$ de longitud. La información de dicha estación meteorológica está disponible al público usuario, a través de consultas en Internet en la página web del Senamhi (<http://www.senamhi.gob.pe/>). Dichos datos consisten de una serie pluviométrica del distrito de Chaupimarca, registrada día a día, donde se midió la cantidad de precipitación habida durante un periodo de 5 años (2009-2013).

Como ya se mencionó en el capítulo 2 de esta tesis, las cadenas de Markov vienen a ser un caso particular de un proceso estocástico de variables aleatorias. Para el estudio que se va a plantear, consideraremos al tiempo y al estado como discretos, usando como unidad de tiempo el día y como estados posibles lluvioso y seco, dependiendo de si se recogió precipitación considerable o no durante el transcurso del día.

Para la medición de la precipitación se utiliza el pluviómetro, que es un instrumento que se emplea en los centros de investigación meteorológica para la recogida y medición de la lluvia caída.

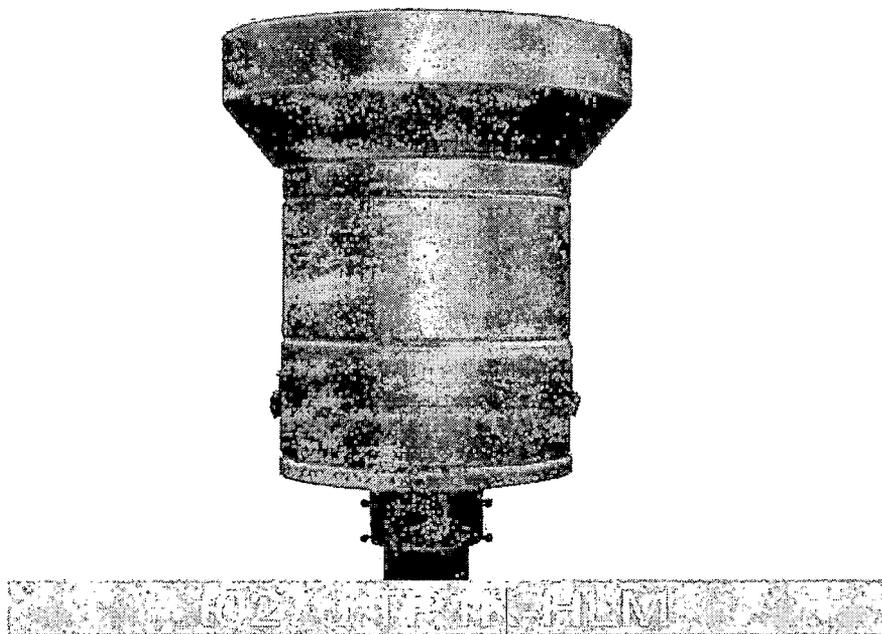


Figura 4.1: *Pluviómetro tipo Hellman modelo t0270 TP1K-HLM*

Se compone de un recipiente cilíndrico, abierto y con el eje vertical, que termina por su parte superior en un borde de latón de filo cortante. El cilindro termina por abajo en una especie de embudo cónico, que en su extremidad inferior lleva una espita (tubo corto que se abre o cierra por el giro de una llave o mediante una palanca); al abrir esta, la lluvia recogida durante un determinado periodo, se transvasa a recipientes graduados. Conociendo la superficie de la base circular del cilindro se obtiene la cantidad de lluvia caída por unidad de superficie en el terreno de la zona. Dicha cantidad se expresa en milímetros (mm), que representan la altura de la capa de agua caída. La dimensión normal de la superficie anteriormente citada en estos instrumentos es de $0,1 \text{ m}^2$, por lo que un litro de agua recogida en el recipiente (equivale a 1 dm^3) representa 10 mm de lluvia.

Ahora que ya conocemos el funcionamiento de un pluviómetro, debemos escoger un umbral bajo el cual un día es considerado seco. En un estudio de precipitación diaria en Badajoz (España) hecho por Cadaña ([29]), se consideró como día seco a todo aquel cuya precipitación

haya sido inferior a 0,1 mm. y como día lluvioso en caso contrario. Otro estudio realizado en el Municipio La Cañada de Urdaneta del Estado de Zulia en Venezuela ([9]), se consideró también como umbral el valor de 0,1 mm. Liao *et al.* ([18]) realizaron un estudio modelación de precipitación diaria en China, donde consideran a un día lluvioso cuando la precipitación es superior a 0,1 mm y a un día seco cuando es menor a dicho valor. De acuerdo a Luengo *et al.* ([21]), se consideraron dos umbrales para determinar los días secos (menor que 0,1 mm y menor que 10 mm), para un estudio sobre rachas secas en la cuenca del Duero (España). Martínez *et al.* ([22]) consideran un umbral de 0,2 mm para los días secos cuando el área escogida para el estudio no presenta precipitación elevada. En un informe presentado por el profesor de la Universidad Nacional de Ingeniería, Edson Plasencia, y al cual se puede acceder mediante el link <http://www.edsonplasencia.com/>, sobre el cambio climático en Cerro de Pasco, se observa que el nivel de precipitación ha disminuido en los últimos años y continúa descendiendo. Es por ello que, en vista de los antecedentes, en este trabajo se utilizó para el análisis el umbral 0,1 mm, por lo tanto, para determinar un día como seco, la precipitación diaria debe ser inferior a 0,1 mm, y será lluvioso en caso contrario.

En base a estudios anteriores como el presentado por Liao ([18]), que considera una cadena de Markov de dos estados (lluvioso y seco); el de Gregory ([11]), que plantea también un modelo de cadena de Markov con diferentes órdenes; el propuesto por Cazacioc ([5]) donde, como en el caso anterior, se plantea dicho modelo para varios órdenes; o en el estudio presentado por Barrantes ([3]), acerca de la persistencia de lluvias en San José (Costa Rica), donde también se plantea este modelo con hasta dos órdenes; vamos a tomar el mismo modelo para nuestro estudio, es decir, el de una cadena de Markov de dos estados, donde el orden lo vamos a determinar en base al criterio de información de Akaike y el criterio de información bayesiano, cuyas metodologías a continuación pasaremos a detallar.

4.3.1. Determinación del orden de una cadena de Markov de dos estados

Para determinar dicho orden, nos basaremos en los dos criterios mencionados anteriormente: el criterio de información de Akaike (*AIC*) y el criterio de información bayesiano (*BIC*). Ambos se basan en las funciones de logverosimilitud de las probabilidades de transición de la cadena de Markov previamente construida. Dichas funciones dependen de las transiciones de un estado a otro y de las probabilidades de transición. Las funciones de logverosimilitud para las cadenas de Markov de orden 0, 1, 2 y 3 se definen como

$$\begin{aligned}
L_0 &= \sum_{j=1}^2 n_j \log(\widehat{p}_j) \\
L_1 &= \sum_{i=1}^2 \sum_{j=1}^2 n_{ij} \log(\widehat{p}_{ij}) \\
L_2 &= \sum_{k=1}^2 \sum_{i=1}^2 \sum_{j=1}^2 n_{kij} \log(\widehat{p}_{kij}) \\
L_3 &= \sum_{l=1}^2 \sum_{k=1}^2 \sum_{i=1}^2 \sum_{j=1}^2 n_{lkij} \log(\widehat{p}_{lkij})
\end{aligned}$$

donde

$$\begin{aligned}
p_{ij} &= P(X_{n+1} = j | X_n = i) \quad \text{y} \quad \widehat{p}_{ij} = \frac{n_{ij}}{n_{i.}} \\
p_{kij} &= P(X_{n+1} = j | X_n = i, X_{n-1} = k) \quad \text{y} \quad \widehat{p}_{kij} = \frac{n_{kij}}{n_{ki.}} \\
p_{lkij} &= P(X_{n+1} = j | X_n = i, X_{n-1} = k, X_{n-2} = l) \quad \text{y} \quad \widehat{p}_{lkij} = \frac{n_{lkij}}{n_{lki.}}
\end{aligned}$$

además

$$n_{i.} = \sum_j n_{ij}, \quad n_{ki.} = \sum_j n_{kij} \quad \text{y} \quad n_{lki.} = \sum_j n_{lkij}$$

Los estadísticos para los dos criterios mencionados son los siguientes:

$$\begin{aligned}
AIC(m) &= -2L_m + 2^{m+1} \\
BIC(m) &= -2L_m + 2^m(\log n)
\end{aligned}$$

donde n representa el tamaño de la muestra. Según Inga ([14]), el criterio que propuso Akaike se basa en calcular el $AIC(m)$ y luego seleccionar el modelo que tenga el mínimo $AIC(m)$. Cazacioc ([5]) afirma que el criterio BIC tiende a ser más conservativo, indicando órdenes más bajos que el criterio AIC . Es preferible usar el criterio BIC para series de tiempo grandes con un número de datos de alrededor de $n = 100$ o superior a $n = 1000$.

Para el cálculo de los valores del $AIC(m)$ y $BIC(m)$ se realizó un programa con el software R, y cuya código se encuentra en el apéndice de este trabajo. Los resultados que se obtuvieron se muestran en la tabla 4.4.

m	AIC	BIC
0	2009,365	2014,875
1	1919,648	1930,668
2	2628,359	2650,399
3	3757,206	3801,285

Tabla 4.4: Valores del AIC y BIC para diferentes órdenes

De acuerdo al criterio de Akaike vemos que el valor mínimo se da cuando $m = 1$, es decir, cuando la cadena de Markov de dos estados es de primer orden (existe una dependencia con un único estado pasado). Para el criterio bayesiano vemos que el valor mínimo se logra también cuando $m = 1$; así, la cadena de Markov de dos estados para el conjunto de datos de precipitación diaria sería de orden uno.

Una vez que ya hemos determinado el orden de la cadena de Markov de dos estados para nuestro conjunto de datos, pasaremos a realizar la inferencia mediante verosimilitud para la cadena no paramétrica.

4.3.2. Simulación de una cadena de Markov no paramétrica

Con los mismos algoritmos para el caso paramétrico haremos simulaciones de una cadena de Markov no paramétrica finita y luego realizaremos la inferencia mediante verosimilitud verificando las probabilidades de cobertura asociadas a los intervalos de verosimilitud.

A partir de la ecuación (4.10) y utilizando la matriz de frecuencias F , la matriz de probabilidades de transición estimada \hat{P} sería

$$\hat{P} = (\hat{p}_{ij}) = \begin{bmatrix} \frac{n_{11}}{n_1} & \frac{n_{12}}{n_1} \\ \frac{n_{21}}{n_2} & \frac{n_{22}}{n_2} \end{bmatrix}$$

$$\hat{P} = (\hat{p}_{ij}) = \begin{bmatrix} 0,817 & 0,183 \\ 0,585 & 0,415 \end{bmatrix} \quad (4.13)$$

Ahora, para hallar el intervalo de verosimilitud para \hat{p}_{11} , utilizamos la función de verosimilitud relativa dada en (4.12)

$$R(p_{11}; x) = \frac{L(p_{11}; x)}{L(\hat{p}_{11}; x)} = \left(\frac{p_{11}}{\hat{p}_{11}} \right)^{n_{11}} \left(\frac{1 - p_{11}}{1 - \hat{p}_{11}} \right)^{n_1 - n_{11}}$$

Haciendo lo mismo para \hat{p}_{12} , \hat{p}_{21} y \hat{p}_{22} , obtenemos ecuaciones similares a la anterior.

$$R(p_{11}; x) = \frac{L(p_{12}; x)}{L(\hat{p}_{12}; x)} = \left(\frac{p_{12}}{\hat{p}_{12}} \right)^{n_{12}} \left(\frac{1 - p_{12}}{1 - \hat{p}_{12}} \right)^{n_1 - n_{12}}$$

$$R(p_{21}; x) = \frac{L(p_{21}; x)}{L(\widehat{p}_{21}; x)} = \left(\frac{p_{21}}{\widehat{p}_{21}} \right)^{n_{21}} \left(\frac{1 - p_{21}}{1 - \widehat{p}_{21}} \right)^{n_2 - n_{21}}$$

$$R(p_{22}; x) = \frac{L(p_{22}; x)}{L(\widehat{p}_{22}; x)} = \left(\frac{p_{22}}{\widehat{p}_{22}} \right)^{n_{22}} \left(\frac{1 - p_{22}}{1 - \widehat{p}_{22}} \right)^{n_2 - n_{22}}$$

Los resultados de los intervalos obtenidos para cada \widehat{p}_{ij} a los diferentes niveles de plausibilidad ($c = 0,036, 0,15$ y $0,25$) se muestran a continuación en las siguientes tablas:

c	A	\widehat{p}_{11}	B
0,036	0,790	0,817	0,843
0,15	0,797	0,817	0,837
0,25	0,800	0,817	0,834

Tabla 4.5: Intervalos de verosimilitud para \widehat{p}_{11}

c	A	\widehat{p}_{12}	B
0,036	0,157	0,183	0,210
0,15	0,163	0,183	0,203
0,25	0,166	0,183	0,200

Tabla 4.6: Intervalos de verosimilitud para \widehat{p}_{12}

c	A	\widehat{p}_{21}	B
0,036	0,524	0,585	0,645
0,15	0,539	0,585	0,631
0,25	0,546	0,585	0,624

Tabla 4.7: Intervalos de verosimilitud para \widehat{p}_{21}

c	A	\widehat{p}_{22}	B
0,036	0,345	0,415	0,476
0,15	0,369	0,415	0,461
0,25	0,376	0,415	0,454

Tabla 4.8: Intervalos de verosimilitud para \widehat{p}_{22}

Realizamos la simulación de una muestra de la cadena de tamaño $n = 1826$ y luego a partir de (4.10) estimaremos las probabilidades de transición \widehat{p}_{ij} . Para ello tomaremos como condición inicial a la matriz de probabilidades de transición dada en (4.13) y, como en el caso paramétrico, consideraremos una distribución estacionaria π_0 que se obtiene al resolver el sistema de ecuaciones dado por

$$\pi_0 \widehat{P} = \pi_0$$

que para este caso sería equivalente a resolver

$$\pi_0(1)\widehat{p}_{11} + \pi_0(2)\widehat{p}_{21} = \pi_0(1)$$

$$\pi_0(1)\widehat{p}_{12} + \pi_0(2)\widehat{p}_{22} = \pi_0(2)$$

con la condición

$$\pi_0(1) + \pi_0(2) = 1$$

La distribución estacionaria π_0 estaría entonces dada por

$$\pi_0(1) = P(X_0 = 1) = 0,762 \text{ y } \pi_0(2) = P(X_0 = 2) = 0,238$$

La nueva matriz de probabilidades de transición obtenida luego de realizar la simulación es

$$\widehat{P} = (\widehat{p}_{ij}) = \begin{bmatrix} 0,815 & 0,184 \\ 0,579 & 0,421 \end{bmatrix}$$

También se hallaron los intervalos de verosimilitud a los diferentes niveles de plausibilidad $c = 0,036, 0,15$ y $0,25$, para las p_{ij} de la matriz anterior (todos estos cálculos fueron hechos mediante un algoritmo implementado con el software R, el cual se mostrará en el apéndice de esta tesis). Los resultados se muestran a continuación en las siguientes tablas:

c	A	\widehat{p}_{11}	B
0,036	0,788	0,815	0,842
0,15	0,795	0,815	0,836
0,25	0,798	0,815	0,833

Tabla 4.9: Intervalos de verosimilitud para \widehat{p}_{11} de la cadena simulada

c	A	\widehat{p}_{12}	B
0,036	0,158	0,184	0,212
0,15	0,164	0,184	0,205
0,25	0,167	0,184	0,202

Tabla 4.10: Intervalos de verosimilitud para \widehat{p}_{12} de la cadena simulada

c	A	\widehat{p}_{21}	B
0,036	0,517	0,579	0,638
0,15	0,532	0,579	0,624
0,25	0,539	0,579	0,617

Tabla 4.11: Intervalos de verosimilitud para \widehat{p}_{21} de la cadena simulada

c	A	\widehat{p}_{22}	B
0,036	0,362	0,421	0,483
0,15	0,376	0,421	0,468
0,25	0,383	0,421	0,461

Tabla 4.12: Intervalos de verosimilitud para \widehat{p}_{22} de la cadena simulada

Ahora verificaremos la confianza asociada a los intervalos de verosimilitud y, al igual que en el caso paramétrico, una vez calculados los intervalos de verosimilitud, verificaremos si el intervalo cubre o no el verdadero valor de p_{ij} (en la primera parte de esta sección, vimos que dichos valores corresponden a los elementos de la matriz de probabilidades de transición estimada \widehat{P} , es decir, $p_{11} = 0,817$, $p_{12} = 0,183$, $p_{21} = 0,585$ y $p_{22} = 0,415$) a los niveles de plausibilidad $c = 0,036$, $0,15$ y $0,25$. Para ello realizaremos el proceso anterior 5000 veces y estimaremos las probabilidades de cobertura (para este proceso nuevamente nos valdremos de un algoritmo implementado con el software R y cuyo código se mostrará en el apéndice de esta tesis).

De la tabla 2.4.4 mostrada abajo vemos que la probabilidades de cobertura (nivel de confianza real) a los diferentes niveles de plausibilidad ($c = 0,036$, $0,15$ y $0,25$) son muy similares a lo teóricamente indicado (nivel de confianza nominal). Al igual que en el caso paramétrico se puede optar por usar otra distribución inicial π_0 , no necesariamente estacionaria, y los resultados de las probabilidades de cobertura serán muy similares.

c	$1 - \alpha$	$PC(p_{11})$	$PC(p_{12})$	$PC(p_{21})$	$PC(p_{22})$
0,036	0,99	0,992	0,992	0,989	0,989
0,15	0,95	0,947	0,947	0,950	0,950
0,25	0,90	0,901	0,901	0,907	0,907

Tabla 4.13: Probabilidades de cobertura a los diferentes niveles de plausibilidad

Si analizamos nuestra cadena representada por la matriz de probabilidades de transición dada en (4.13) desde el enfoque probabilístico, vemos que es *finita* pues tiene dos estados ($T = \{1, 2\}$). Además, como todas sus entradas son positivas ($p_{ij} > 0, \forall i, j \in T$), se cumple que $p_{ij}^{(n)} > 0, \forall i, j \in T$. Así, vemos que todos sus estados son comunicantes y por tanto la cadena sería *irreducible*.

Pasemos ahora a ver la recurrencia de nuestra cadena. Como recordamos, se dice que un estado es recurrente si f_{ii} , donde f_{ii} denota la probabilidad de regresar al estado i en un tiempo positivo. Si queremos conocer el tiempo medio que la cadena tardará en regresar a dicho estado i , debemos analizar la recurrencia positiva. Por ser nuestra cadena finita y de acuerdo a la proposición 2.5.15, entonces se tiene al menos un estado recurrente positivo. Además, como nuestra cadena es irreducible y finita, por el corolario 2.5.5 tendríamos que cada estado es recurrente positivo. Por tanto tenemos que nuestra cadena es *recurrente positiva*.

La periodicidad mide la cantidad de pasos en los cuales un estado regresa así mismo. Como $p_{11} > 0$, entonces $p_{11}^{(2)} > 0$; en general, $p_{11}^{(n)} > 0$. De la definición de periodo tendríamos que $d(1) = 1$ y por la proposición 2.5.1, $d(2) = 1$. Así, nuestra cadena sería *aperiódica*.

Vimos que nuestra distribución estacionaria estaba dada por $\pi_0 = (0,762; 0,238)$. De acuerdo al teorema 2.5.1, como nuestra cadena es irreducible, recurrente positiva, aperiódica, con distribución estacionaria π_0 , entonces para una cantidad grande de pasos ($n = 20$)

$$\widehat{P}^{(20)} = (\widehat{p}_{ij}^{(20)}) = \begin{bmatrix} 0,762 & 0,238 \\ 0,762 & 0,238 \end{bmatrix}$$

las filas de la matriz anterior tienden a la distribución estacionaria π_0 , la cual sería la distribución límite (dicha matriz fue calculada computacionalmente).

Conclusiones

1. Se probó que existe una relación entre la estadística y la probabilidad valiéndonos para ello de la inferencia estadística aplicada sobre las probabilidades de transición de una cadena de Markov paramétrica y no paramétrica. La teoría de probabilidad se aplica al analizar el comportamiento asintótico de dicha cadena.
2. Mediante el método de verosimilitud se mostró cómo estimar dichas probabilidades de transición tanto para el caso paramétrico como para el no paramétrico. En la práctica es más frecuente encontrarse con el segundo caso puesto que el primero es muchas veces muy complicado de modelar.
3. No solo se realizó la estimación puntual, sino también la estimación por intervalos. La propiedad de verosimilitud de cobertura de intervalos fue verificada, mediante un proceso de simulación, para ver si esta cubre o no el verdadero valor del parámetro estimado. Se comparó dicho valor (nivel de confianza real) con el valor teórico (nivel de confianza nominal) viendo que estos valores están muy próximos.
4. Para el ejemplo de aplicación real, sobre la ocurrencia o no ocurrencia de precipitación, nuestra cadena de Markov consiste de dos estados. Según la matriz de probabilidades de transición vemos que si el día de hoy no llovió es más probable que el día de mañana no llueva ($p_{11} = 0,817$), de que el día de mañana llueva ($p_{12} = 0,183$). Asimismo, si el día de hoy llovió es más probable que el día de mañana no llueva ($p_{21} = 0,585$), de que el día de mañana llueva ($p_{22} = 0,415$).
5. La cadena de Markov en el presente trabajó resultó ser finita, irreducible, recurrente positiva y aperiódica. Además, según el teorema básico de convergencia, vemos que las probabilidades de transición de la matriz P^n convergen a la distribución estacionaria π_0 , la cual sería la distribución límite.

Apéndice

1. Código en R del ejemplo 3.6.1

```
# Calculo del EMV para la distribucion exponencial

tf=c(2,51,33,27,14,24,4) # Tiempos de falla
tf
T=c(72,60,21) # Tiempos de vida censurados
T
r=length(tf)
r
t=sum(T)+sum(tf)
t
theta_esti=t/r
theta_esti

# Grafica de R(theta) y los niveles para c=0.25, 0.15 y 0.036

theta=seq(0,180,by=1)
R=(7*theta/308)^(-7)*exp((-308/theta)+7)
plot(theta,R,type="l",xlab=expression(theta),ylab=expression("R(" * theta * ";" x
)"))
legend(130,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","
green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=theta_esti)

# Intervalos de confianza para theta estimado

f1 <- function (x) (7*x/308)^(-7)*exp((-308/x)+7)-0.036
xmin1 <- uniroot(f1, c(-1, theta_esti), tol = 0.0001)$root
xmax1 <- uniroot(f1, lower=theta_esti, upper=150, tol = 0.0001)$root
```

```

I1<-c(xmin1,xmax1)
I1

f2 <- function (x) (7*x/308)^(-7)*exp((-308/x)+7)-0.15
xmin2 <- uniroot(f2, c(-1, theta_esti), tol = 0.0001)$root
xmax2 <- uniroot(f2, lower=theta_esti, upper=150, tol = 0.0001)$root
I2<-c(xmin2,xmax2)
I2

f3 <- function (x) (7*x/308)^(-7)*exp((-308/x)+7)-0.25
xmin3 <- uniroot(f3, c(-1, theta_esti), tol = 0.0001)$root
xmax3 <- uniroot(f3, lower=theta_esti, upper=150, tol = 0.0001)$root
I3<-c(xmin3,xmax3)
I3

```

2. Código en R del ejemplo 4.1.1

```

# Calculo de las frecuencias para dos estados

X=read.table("datos.txt",header=TRUE)
X
n=dim(X)
n
T=c(1,2)
m=length(T)

F=matrix(0,nrow=m,ncol=m) # Matriz de frecuencias
for (i in 1:(n[1]-1))
{
  for (j1 in 1:m)
  {
    if (X[i,2]==1 & X[i+1,2]==j1)
    {
      F[1,j1]=F[1,j1]+1
    }
  }
  for (j2 in 1:m)
  {
    if (X[i,2]==2 & X[i+1,2]==j2)
    {
      F[2,j2]=F[2,j2]+1
    }
  }
}
}

```

```

F

n11=F[1,1]
n12=F[1,2]
n21=F[2,1]
n22=F[2,2]
n1=n11+n12
n2=n21+n22

# Calculo del valor estimado para theta
# Raices del polinomio (n11+n12+6n12+6n22)x^7-(n11+6n21+6n22)x^6-(n11+n12+6
n22)x+(n11+6n22)=0

a=n11+n12+6*n12+6*n22
b=n11+6*n21+6*n22
c=n11+n12+6*n22
d=n11+6*n22

z=polyroot(c(d,-c,0,0,0,0,-b,a))
z
r=Re(z[4])
r

# Grafica de R(theta) y los niveles para c=0.25, 0.15 y 0.036

theta=seq(0.80,0.90,by=0.001)
R=(theta/r)^(d)*((1-theta)/(1-r))^(n12)*((1-theta^6)/(1-r^6))^(n21)
par(mfrow=c(1,1))
plot(theta,R,type="l",xlab="Theta",ylab="Verosimilitud relativa (R)")
legend(0.87,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","
green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=r)

# Derivada de la funcion de verosimilitud relativa

DD <- function(expre, nombre, orden = 1) {
  if(orden < 1) stop("'orden' debe ser >= 1")
  if(orden == 1) D(expre, nombre)
  else DD(D(expre, nombre), nombre, orden - 1)
}
DD(expression((x/r)^(d)*((1-x^3)/(1-r^3))^(n12)*((1-x)/(1-r))^(n21)),"x",

```

```

1) # 1 es el orden de la derivada

x=seq(0.70,0.95,by=0.001)
y=((x/r)^(d - 1) * ((d) * (1/r)) * ((1 - x^3)/(1 - r^3))^(n12) -
  (x/r)^d * (((1 - x^3)/(1 - r^3))^((n12) - 1) * ((n12) *
    (3 * x^2/(1 - r^3)))) * ((1 - x)/(1 - r))^(n21) - (x/r)^d *
  ((1 - x^3)/(1 - r^3))^(n12) * (((1 - x)/(1 - r))^((n21) -
    1) * ((n21) * (1/(1 - r))))
plot(x,y,type="l",xlab="x",ylab="y",panel.first=grid())
abline(h=0,col="blue")

# Intervalos de confianza para theta estimado

f1 <- function (x, a) (x/a)^d*((1-x)/(1-a))^(n12)*((1-x^6)/(1-a^6))^(n21)
-0.036
xmin1 <- uniroot(f1, c(0.80, r), tol = 0.0001, a=r)$root
xmax1 <- uniroot(f1, lower=r, upper=0.92, tol = 0.0001, a=r)$root
I1<-c(xmin1,xmax1)
I1

f2 <- function (x, a) (x/a)^d*((1-x)/(1-a))^(n12)*((1-x^6)/(1-a^6))^(n21)
-0.15
xmin2 <- uniroot(f2, c(0.80, r), tol = 0.0001, a=r)$root
xmax2 <- uniroot(f2, lower=r, upper=0.92, tol = 0.0001, a=r)$root
I2<-c(xmin2,xmax2)
I2

f3 <- function (x, a) (x/a)^d*((1-x)/(1-a))^(n12)*((1-x^6)/(1-a^6))^(n21)
-0.25
xmin3 <- uniroot(f3, c(0.80, r), tol = 0.0001, a=r)$root
xmax3 <- uniroot(f3, lower=r, upper=0.92, tol = 0.0001, a=r)$root
I3<-c(xmin3,xmax3)
I3

P_esti=matrix(c(r,1-r^6,1-r,r^6),nrow=m,ncol=m)
P_esti

```

3. Código en R de la simulación de una cadena de Markov paramétrica

```

# Determinacion de la distribucion inicial pi_0

T=c(1,2)
m=length(T)
r=0.843 # EMV del ejemplo 6.1.1

```

```

P=matrix(c(r,1-r^6,1-r,r^6),nrow=m,ncol=m) # Matriz P_esti del ejemplo
6.1.1
P
A=matrix(c(P[1,1]-1,P[1,2]+1,P[2,1],P[2,2]),nrow=m,ncol=m)
A
b=c(0,1)
b
pi_0=solve(A,b) # Distribucion estacionaria pi_0 del ejemplo 6.1.1
pi_0

# Simulacion de una cadena de Markov en {1, 2} dadas una
# condicion inicial pi_0 y matriz de transicion P

N=1826 # Numero de datos a simular (N=Total de datos+1)
X=matrix(0,nrow=1,ncol=N+1) # Inicializar y borrar si hay datos anteriores
t=0:N # Indices de tiempo

# rando genera numeros aleatorios (1 o 2) con distribucion Uniforme([0,1])

rando<-function(p)
{
u=runif(1,min=0,max=1)
s=p[1]
j=1

while ((u>s) & (j<length(p)))
{
j=j+1
s=s+p[j]
}

index=j
}

X[1]=rando(pi_0) # Genera el valor de X(0) con distribucion pi_0

for (i in 1:N)
{
X[i+1]=rando(P[X[i],])
}
X

par(mfrow=c(2,1))

plot(t,X,type="b",col='4',pch=18,main="Una trayectoria de la cadena")

```

```

hist(X,col='4',breaks=seq(1,2,by=1/2),ylim=c(0,2000),main="Frecuencias
relativas de la cantidad de visitas a cada estado")

n=length(X)
n

F=matrix(0,nrow=m,ncol=m) # Matriz de frecuencias
for (i in 1:(n-1))
{
  for (j1 in 1:m)
  {
    if (X[i]==1 & X[i+1]==j1)
    {
      F[1,j1]=F[1,j1]+1
    }
  }
  for (j2 in 1:m)
  {
    if (X[i]==2 & X[i+1]==j2)
    {
      F[2,j2]=F[2,j2]+1
    }
  }
}
F

n11=F[1,1]
n12=F[1,2]
n21=F[2,1]
n22=F[2,2]

# Calculo del valor estimado para theta
# Raices del polinomio (n11+n12+6n21+6n22)x^7-(n11+6n21+6n22)x^6-(n11+n12+6
n22)x+(n11+6n22)=0

a=n11+n12+6*n21+6*n22
b=n11+6*n21+6*n22
c=n11+n12+6*n22
d=n11+6*n22

z=polyroot(c(d,-c,0,0,0,0,-b,a))
z
r.sim=Re(z[4])
r.sim

```

```

# Grafica de R(theta) y los niveles para c=0.25, 0.15 y 0.036

theta=seq(0.8,0.9,by=0.001)
R=(theta/r_sim)^(d)*((1-theta)/(1-r_sim))^(n12)*((1-theta^6)/(1-r_sim^6))^(
n21)
par(mfrow=c(1,1))
plot(theta,R,type="l",xlab=expression(theta),ylab=expression("R"*theta*";x
"))
legend(0.87,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","
green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=r_sim)

# Derivada de la funcion de verosimilitud relativa

DD <- function(expre, nombre, orden = 1) {
  if(orden < 1) stop("'orden' debe ser >= 1")
  if(orden == 1) D(expre, nombre)
  else DD(D(expre, nombre), nombre, orden - 1)
}
DD(expression((x/r_sim)^(d)*((1-x^3)/(1-r_sim^3))^(n12)*((1-x)/(1-r_sim))^(
n21)), "x", 1) # 1 es el orden de la derivada

x=seq(0.70,0.95,by=0.001)
y=((x/r_sim)^((d) - 1) * ((d) * (1/r_sim)) * ((1 - x^3)/(1 - r_sim^3))^(n12
) -
(x/r_sim)^(d) * (((1 - x^3)/(1 - r_sim^3))^(n12) - 1) *
((n12) * (3 * x^2/(1 - r_sim^3)))) * ((1 - x)/(1 - r_sim))^(n21) -
(x/r_sim)^(d) * ((1 - x^3)/(1 - r_sim^3))^(n12) * (((1 -
x)/(1 - r_sim))^(n21) - 1) * ((n21) * (1/(1 - r_sim))))

plot(x,y,type="l",xlab="x",ylab="y",panel.first=grid())
abline(h=0,col="blue")

# Intervalos de confianza para theta estimado

f1 <- function(x, a) (x/a)^(d)*((1-x)/(1-a))^(n12)*((1-x^6)/(1-a^6))^(n21)
-0.036
xmin1 <- uniroot(f1, c(0.75, r_sim), tol = 0.0001, a=r_sim)$root
xmax1 <- uniroot(f1, lower=r_sim, upper=0.92, tol = 0.0001, a=r_sim)$root
I1<-c(xmin1,xmax1)

```

```

I1

f2 <- function (x, a) (x/a)^(d)*((1-x)/(1-a))^(n12)*((1-x^6)/(1-a^6))^(n21)
-0.15
xmin2 <- uniroot(f2, c(0.75, r_sim), tol = 0.0001, a=r_sim)$root
xmax2 <- uniroot(f2, lower=r_sim, upper=0.92, tol = 0.0001, a=r_sim)$root
I2<-c(xmin2,xmax2)
I2

f3 <- function (x, a) (x/a)^(d)*((1-x)/(1-a))^(n12)*((1-x^6)/(1-a^6))^(n21)
-0.25
xmin3 <- uniroot(f3, c(0.75, r_sim), tol = 0.0001, a=r_sim)$root
xmax3 <- uniroot(f3, lower=r_sim, upper=0.92, tol = 0.0001, a=r_sim)$root
I3<-c(xmin3,xmax3)
I3

P_esti=matrix(c(r_sim,1-r_sim^6,1-r_sim,r_sim^6),nrow=m,ncol=n)
P_esti

```

4. Código en R para calcular la probabilidad de cobertura para el caso paramétrico

```

theta_esti=0.843
S=5000
k=1
cont1=0
cont2=0
cont3=0
N=1826
xmin1=rep(0,N)
xmax1=rep(0,N)
xmin2=rep(0,N)
xmax2=rep(0,N)
xmin3=rep(0,N)
xmax3=rep(0,N)

while (k<S)
{
# Simulacion de una cadena de Markov en {0, 1} dadas una
# condicion inicial pi_0 y matriz de transicion P

pi_0=c(0.803,0.197)
pi_0          # Distribucion inicial
P=matrix(c(0.843,0.641,0.157,0.359),nrow=2,ncol=2) # Matriz de

```

```

    probabilidades de transicion
P
N=1826          # Numero de pasos a simular (N=Total de datos+1)
X=matrix(0,nrow=1,ncol=N+1) # Inicializar y borrar si hay datos anteriores
t=0:N          # Indices de tiempo

# rando genera numeros aleatorios (1 o 2) con distribucion Uniforme([0,1])

rando<-function(p)
{
u=runif(1,min=0,max=1)
s=p[1]
j=1

while ((u>s) & (j<length(p)))
{
j=j+1
s=s+p[j]
}

index=j
}

X[1]=rando(pi.0) # Genera el valor de X(0) con distribucion pi.0

for (i in 1:N)
{
X[i+1]=rando(P[X[i],])
}

# Calculo de las frecuencias para dos estados

n11=0
n12=0
n21=0
n22=0
for (i in 1:(N-1))
{
if (X[i]==1 & X[i+1]==1)
{
n11=n11+1
}
if (X[i]==1 & X[i+1]==2)
{
n12=n12+1
}
}

```

```

    }
    if (X[i]==2 & X[i+1]==1)
    {
        n21=n21+1
    }
    if (X[i]==2 & X[i+1]==2)
    {
        n22=n22+1
    }
}
n11
n12
n21
n22
F=matrix(c(n11, n21, n12, n22), nrow=2, ncol=2)
F

# Calculo del valor estimado para theta
# Raices del polinomio (n11+n12+6n21+6n22)x^7-(n11+6n21+6n22)x^6-(n11+n12+6
n22)x+(n11+6n22)=0

a=n11+n12+6*n21+6*n22
b=n11+6*n21+6*n22
c=n11+n12+6*n22
d=n11+6*n22

z=polyroot(c(d,-c,0,0,0,0,-b,a))
#z
r_sim=Re(z[4])
#r_sim

# Grafica de R(theta) y los niveles para c=0.25, 0.15 y 0.036

theta=seq(0.8,0.9,by=0.001)
R=(theta/r_sim)^(d)*((1-theta)/(1-r_sim))^(n12)*((1-theta^6)/(1-r_sim^6))^(
n21)
par(mfrow=c(1,1))
plot(theta,R,type="l",xlab=expression(theta),ylab=expression("R("*theta*";x
)"))
legend(0.87,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","
green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")

```

```

abline(v=r.sim)

# Intervalos de confianza para theta estimado

f1 <- function (x, a) (x/a)^(d)*((1-x)/(1-a))^(n12)*((1-x^6)/(1-a^6))^(n21)
-0.036
xmin1[k] <- uniroot(f1, c(0.75, r.sim), tol = 0.0001, a=r.sim)$root
xmax1[k] <- uniroot(f1, lower=r.sim, upper=0.92, tol = 0.0001, a=r.sim)
  $root

f2 <- function (x, a) (x/a)^(d)*((1-x)/(1-a))^(n12)*((1-x^6)/(1-a^6))^(n21)
-0.15
xmin2[k] <- uniroot(f2, c(0.75, r.sim), tol = 0.0001, a=r.sim)$root
xmax2[k] <- uniroot(f2, lower=r.sim, upper=0.92, tol = 0.0001, a=r.sim)
  $root

f3 <- function (x, a) (x/a)^(d)*((1-x)/(1-a))^(n12)*((1-x^6)/(1-a^6))^(n21)
-0.25
xmin3[k] <- uniroot(f3, c(0.75, r.sim), tol = 0.0001, a=r.sim)$root
xmax3[k] <- uniroot(f3, lower=r.sim, upper=0.92, tol = 0.0001, a=r.sim)
  $root

if (theta_esti>xmin1[k] & theta_esti<xmax1[k])
{
  cont1=cont1+1
}
if (theta_esti>xmin2[k] & theta_esti<xmax2[k])
{
  cont2=cont2+1
}
if (theta_esti>xmin3[k] & theta_esti<xmax3[k])
{
  cont3=cont3+1
}
k=k+1
}

# Probabilidades de cobertura a los niveles c=0.036, 0.15 y 0.25

PC1=cont1/S
PC1
PC2=cont2/S
PC2
PC3=cont3/S

```

5. Código en R del ejemplo de aplicación

```

# Calculo de las frecuencias para dos estados

X=read.table("datos.txt",header=T)
X
n=dim(X)
n
T=c(1,2)
m=length(T)

F=matrix(0,nrow=m,ncol=m) # Matriz de frecuencias
for (i in 1:(n[1]-1))
{
  for (j1 in 1:m)
  {
    if (X[i,2]==1 & X[i+1,2]==j1)
    {
      F[1,j1]=F[1,j1]+1
    }
  }
  for (j2 in 1:m)
  {
    if (X[i,2]==2 & X[i+1,2]==j2)
    {
      F[2,j2]=F[2,j2]+1
    }
  }
}
F

n11=F[1,1]
n12=F[1,2]
n21=F[2,1]
n22=F[2,2]
n1=n11+n12
n2=n21+n22

P_esti=matrix(0,nrow=m,ncol=m) # Matriz de probabilidades de transicion
for (i in 1:m)
{
  for (j in 1:m)

```

```

    {
      P_esti[j,i]=F[j,i]/sum(F[j,])
    }
  }
P_esti

# Grafica de R(P[1,1]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.70,0.90,by=0.001)
R=(p/P_esti[1,1])^(n11)*((1-p)/(1-P_esti[1,1]))^(n1-n11)
par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[11]),ylab=expression("R"*p[11]*" ; x"))
legend(0.72,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","
green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[1,1])

# Intervalos de confianza para P[1,1]

f1 <- function (x, a) (x/a)^(n11)*((1-x)/(1-a))^(n1-n11)-0.036
xmin1 <- uniroot(f1, c(0.75, P_esti[1,1]), tol = 0.0001, a=P_esti[1,1])
  $root
xmax1 <- uniroot(f1, lower=P_esti[1,1], upper=0.90, tol = 0.0001, a=P_esti
  [1,1])$root
I1<-c(xmin1,xmax1)
I1

f2 <- function (x, a) (x/a)^(n11)*((1-x)/(1-a))^(n1-n11)-0.15
xmin2 <- uniroot(f2, c(0.75, P_esti[1,1]), tol = 0.0001, a=P_esti[1,1])
  $root
xmax2 <- uniroot(f2, lower=P_esti[1,1], upper=0.90, tol = 0.0001, a=P_esti
  [1,1])$root
I2<-c(xmin2,xmax2)
I2

f3 <- function (x, a) (x/a)^(n11)*((1-x)/(1-a))^(n1-n11)-0.25
xmin3 <- uniroot(f3, c(0.75, P_esti[1,1]), tol = 0.0001, a=P_esti[1,1])
  $root
xmax3 <- uniroot(f3, lower=P_esti[1,1], upper=0.90, tol = 0.0001, a=P_esti
  [1,1])$root
I3<-c(xmin3,xmax3)
I3

```

```

# Grafica de R(P[1,2]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.10,0.30,by=0.001)
R=(p/P_esti[1,2])^(n12)*((1-p)/(1-P_esti[1,2]))^(n1-n12)
par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[12]),ylab=expression("R"*p[12]*";x"))
legend(0.25,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[1,2])

# Intervalos de confianza para P[1,2]

f1 <- function (x, a) (x/a)^(n12)*((1-x)/(1-a))^(n1-n12)-0.036
xmin1 <- uniroot(f1, c(0.1, P_esti[1,2]), tol = 0.0001, a=P_esti[1,2])$root
xmax1 <- uniroot(f1, lower=P_esti[1,2], upper=0.25, tol = 0.0001, a=P_esti[1,2])$root
I1<-c(xmin1,xmax1)
I1

f2 <- function (x, a) (x/a)^(n12)*((1-x)/(1-a))^(n1-n12)-0.15
xmin2 <- uniroot(f2, c(0.1, P_esti[1,2]), tol = 0.0001, a=P_esti[1,2])$root
xmax2 <- uniroot(f2, lower=P_esti[1,2], upper=0.25, tol = 0.0001, a=P_esti[1,2])$root
I2<-c(xmin2,xmax2)
I2

f3 <- function (x, a) (x/a)^(n12)*((1-x)/(1-a))^(n1-n12)-0.25
xmin3 <- uniroot(f3, c(0.1, P_esti[1,2]), tol = 0.0001, a=P_esti[1,2])$root
xmax3 <- uniroot(f3, lower=P_esti[1,2], upper=0.25, tol = 0.0001, a=P_esti[1,2])$root
I3<-c(xmin3,xmax3)
I3

# Grafica de R(P[2,1]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.4,0.7,by=0.001)
R=(p/P_esti[2,1])^(n21)*((1-p)/(1-P_esti[2,1]))^(n2-n21)
par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[21]),ylab=expression("R"*p[21]*";x"))
legend(0.41,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","green"),lty=1)

```

```

    green"), lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[2,1])

# Intervalos de confianza para P[2,1]

f1 <- function (x, a) (x/a)^(n21)*((1-x)/(1-a))^(n2-n21)-0.036
xmin1 <- uniroot(f1, c(0.5, P_esti[2,1]), tol = 0.0001, a=P_esti[2,1])$root
xmax1 <- uniroot(f1, lower=P_esti[2,1], upper=0.65, tol = 0.0001, a=P_esti
  [2,1])$root
I1<-c(xmin1,xmax1)
I1

f2 <- function (x, a) (x/a)^(n21)*((1-x)/(1-a))^(n2-n21)-0.15
xmin2 <- uniroot(f2, c(0.5, P_esti[2,1]), tol = 0.0001, a=P_esti[2,1])$root
xmax2 <- uniroot(f2, lower=P_esti[2,1], upper=0.65, tol = 0.0001, a=P_esti
  [2,1])$root
I2<-c(xmin2,xmax2)
I2

f3 <- function (x, a) (x/a)^(n21)*((1-x)/(1-a))^(n2-n21)-0.25
xmin3 <- uniroot(f3, c(0.5, P_esti[2,1]), tol = 0.0001, a=P_esti[2,1])$root
xmax3 <- uniroot(f3, lower=P_esti[2,1], upper=0.65, tol = 0.0001, a=P_esti
  [2,1])$root
I3<-c(xmin3,xmax3)
I3

# Grafica de R(P[2,2]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.25,0.65,by=0.001)
R=(p/P_esti[2,2])^(n22)*((1-p)/(1-P_esti[2,2]))^(n2-n22)
par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[22]),ylab=expression("R(" * p[22] * "; x)"))
legend(0.51,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","
  green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[2,2])

# Intervalos de confianza para P[2,2]

```

```

f1 <- function (x, a) (x/a)^(n22)*((1-x)/(1-a))^(n2-n22)-0.036
xmin1 <- uniroot(f1, c(0.35, P_esti[2,2]), tol = 0.0001, a=P_esti[2,2])
  $root
xmax1 <- uniroot(f1, lower=P_esti[2,2], upper=0.6, tol = 0.0001, a=P_esti
  [2,2])$root
I1<-c(xmin1,xmax1)
I1

f2 <- function (x, a) (x/a)^(n22)*((1-x)/(1-a))^(n2-n22)-0.15
xmin2 <- uniroot(f2, c(0.35, P_esti[2,2]), tol = 0.0001, a=P_esti[2,2])
  $root
xmax2 <- uniroot(f2, lower=P_esti[2,2], upper=0.6, tol = 0.0001, a=P_esti
  [2,2])$root
I2<-c(xmin2,xmax2)
I2

f3 <- function (x, a) (x/a)^(n22)*((1-x)/(1-a))^(n2-n22)-0.25
xmin3 <- uniroot(f3, c(0.35, P_esti[2,2]), tol = 0.0001, a=P_esti[2,2])
  $root
xmax3 <- uniroot(f3, lower=P_esti[2,2], upper=0.6, tol = 0.0001, a=P_esti
  [2,2])$root
I3<-c(xmin3,xmax3)
I3

# Multiplicacion de la matriz P_esti consigo misma 20 veces

Pn=diag(2)
for (i in 1:20)
{
  Pn=Pn%*%(P_esti)
}
t(Pn)

```

6. Código en R para determinar el orden la cadena de Markov no parmétrica de dos estados

```

# Codigo en R para el criterio de Akaike y bayesiano

X=read.table("datos.txt",header=TRUE)
X
n=dim(X)
n
T=c(1,2)

```

```

n=length(T)

# Calculo de la logverosimilitud de la cadena de Markov de orden 0

F0=matrix(0,nrow=1,ncol=m) # Matriz de frecuencias
for (i in 1:(n[1]))
{
  if (X[i,2]==1)
  {
    F0[1,1]=F0[1,1]+1
  }
  if (X[i,2]==2)
  {
    F0[1,2]=F0[1,2]+1
  }
}
F0
P0=F0/(sum(F0))
P0
L0=0
for (i in 1:m)
{
  L0=L0+F0[1,i]*log(P0[1,i])
}
L0

# Calculo de la logverosimilitud de la cadena de Markov de orden 1

F1=matrix(0,nrow=m,ncol=m) # Matriz de frecuencias
for (i in 1:(n[1]-1))
{
  for (j1 in 1:m)
  {
    if (X[i,2]==1 & X[i+1,2]==j1)
    {
      F1[1,j1]=F1[1,j1]+1
    }
  }
  for (j2 in 1:m)
  {
    if (X[i,2]==2 & X[i+1,2]==j2)
    {
      F1[2,j2]=F1[2,j2]+1
    }
  }
}

```

```

}
F1
P1=matrix(0,nrow=n,ncol=n)
for (i in 1:m)
{
  for (j in 1:m)
  {
    P1[i,j]=F1[i,j]/sum(F1[i,])
  }
}
P1
L1=sum(F1*log(P1)) # Multiplicacion elemento a elemento
L1

# Calculo de la logverosimilitud de la cadena de Markov de orden 2

F2=matrix(0,nrow=4,ncol=n) # Matriz de frecuencias
for (i in 1:(n[1]-2))
{
  for (j1 in 1:m)
  {
    if (X[i,2]==1 & X[i+1,2]==1 & X[i+2,2]==j1)
    {
      F2[1,j1]=F2[1,j1]+1
    }
  }
  for (j2 in 1:m)
  {
    if (X[i,2]==1 & X[i+1,2]==2 & X[i+2,2]==j2)
    {
      F2[2,j2]=F2[2,j2]+1
    }
  }
  for (j3 in 1:m)
  {
    if (X[i,2]==2 & X[i+1,2]==1 & X[i+2,2]==j3)
    {
      F2[3,j3]=F2[1,j3]+1
    }
  }
  for (j4 in 1:m)
  {
    if (X[i,2]==2 & X[i+1,2]==2 & X[i+2,2]==j4)
    {
      F2[4,j4]=F2[2,j4]+1
    }
  }
}

```

```

    }
  }
}
F2
P2=matrix(0,nrow=4,ncol=m)
for (i in 1:4)
{
  for (j in 1:m)
  {
    P2[i,j]=F2[i,j]/sum(F2[i,])
  }
}
P2
L2=sum(F2*log(P2)) # Multiplicacion elemento a elemento
L2

# Calculo de la logverosimilitud de la cadena de Markov de orden 3

F3=matrix(0,nrow=4,ncol=4) # Matriz de frecuencias
for (i in 1:(n[1]-3))
{
  for (j1 in 1:m)
  {
    if (X[i,2]==1 & X[i+1,2]==1 & X[i+2,2]==1 & X[i+3,2]==j1)
    {
      F3[1,j1]=F3[1,j1]+1
    }
    if (X[i,2]==1 & X[i+1,2]==1 & X[i+2,2]==2 & X[i+3,2]==j1)
    {
      F3[1,j1+2]=F3[1,j1+2]+1
    }
  }
  for (j2 in 1:m)
  {
    if (X[i,2]==1 & X[i+1,2]==2 & X[i+2,2]==1 & X[i+3,2]==j2)
    {
      F3[2,j2]=F3[2,j2]+1
    }
    if (X[i,2]==1 & X[i+1,2]==2 & X[i+2,2]==2 & X[i+3,2]==j2)
    {
      F3[2,j2+2]=F3[2,j2+2]+1
    }
  }
  for (j3 in 1:m)
  {

```

```

    if (X[i,2]==2 & X[i+1,2]==1 & X[i+2,2]==1 & X[i+3,2]==j3)
    {
        F3[3,j3]=F3[3,j3]+1
    }
    if (X[i,2]==2 & X[i+1,2]==1 & X[i+2,2]==2 & X[i+3,2]==j3)
    {
        F3[3,j3+2]=F3[3,j3+2]+1
    }
}
for (j4 in 1:m)
{
    if (X[i,2]==2 & X[i+1,2]==2 & X[i+2,2]==1 & X[i+3,2]==j4)
    {
        F3[4,j4]=F3[4,j4]+1
    }
    if (X[i,2]==2 & X[i+1,2]==2 & X[i+2,2]==2 & X[i+3,2]==j4)
    {
        F3[4,j4+2]=F3[4,j4+2]+1
    }
}
}
F3
P3=matrix(0,nrow=4,ncol=4)
for (i in 1:4)
{
    for (j in 1:4)
    {
        P3[i,j]=F3[i,j]/sum(F3[i,])
    }
}
P3
L3=sum(F3*log(P3)) # Multiplicacion elemento a elemento
L3
AIC=c(-2*L0+2*(m-1),-2*L1+2*m*(m-1),-2*L2+2*m^2*(m-1),-2*L3+2*m^3*(m-1))
AIC
BIC=c(-2*L0+log(n[1]),-2*L1+m*log(n[1]),-2*L2+m^2*log(n[1]),-2*L3+m^3*log(n
[1]))
BIC

```

7. Código en R de la simulación de una cadena de Markov no paramétrica

```
# Calculo de las frecuencias para dos estados
```

```

X=read.table("datos.txt",header=F)
X
n=dim(X)
n
T=c(1,2)
m=length(T)

F=matrix(0,nrow=m,ncol=m) # Matriz de frecuencias
for (i in 1:(n[1]-1))
{
  for (j1 in 1:m)
  {
    if (X[i,2]==1 & X[i+1,2]==j1)
    {
      F[1,j1]=F[1,j1]+1
    }
  }
  for (j2 in 1:m)
  {
    if (X[i,2]==2 & X[i+1,2]==j2)
    {
      F[2,j2]=F[2,j2]+1
    }
  }
}
F

n11=F[1,1]
n12=F[1,2]
n21=F[2,1]
n22=F[2,2]
n1=n11+n12
n2=n21+n22

P=matrix(0,nrow=m,ncol=m) # Matriz de probabilidades de transicion
for (i in 1:m)
{
  for (j in 1:m)
  {
    P[j,i]=F[j,i]/sum(F[j,])
  }
}
P

```

```

A=matrix(c(P[1,1]-1,P[1,2]+1,P[2,1],P[2,2]),nrow=m,ncol=m)
A
b=c(0,1)
b
pi_0=solve(A,b) # Distribucion estacionaria pi_0 del ejemplo 6.1.1
pi_0

# Simulacion de una cadena de Markov en {1, 2} dadas una
# condicion inicial pi_0 y una matriz de transicion P

N=sum(F) # Numero de datos a simular (N=Total de datos+1)
X=matrix(0,nrow=1,ncol=N+1) # Inicializar y borrar si hay datos anteriores
t=0:N # Indices de tiempo

# rando genera numeros aleatorios (1 o 2) con distribucion Uniforme([0,1])

rando<-function(p)
{
u=runif(1,min=0,max=1)
s=p[1]
j=1

while ((u>s) & (j<length(p)))
{
j=j+1
s=s+p[j]
}

index=j
}

X[1]=rando(pi_0) # Genera el valor de X(0) con distribucion pi_0

for (i in 1:N)
{
X[i+1]=rando(P[X[i],])
}
X

par(mfrow=c(2,1))

plot(t,X,type="b",col='4',pch=18,main="Una trayectoria de la cadena")

hist(X,col='4',breaks=seq(1,2,by=1/2),ylim=c(0,N),main="Frecuencias
relativas de la cantidad de visitas a cada estado")

```

```

n=length(X)
n
F=matrix(0,nrow=m,ncol=m)
for (i in 1:(n-1))
{
  for (j1 in 1:m)
  {
    if (X[i]==1 & X[i+1]==j1)
    {
      F[1,j1]=F[1,j1]+1
    }
  }
  for (j2 in 1:m)
  {
    if (X[i]==2 & X[i+1]==j2)
    {
      F[2,j2]=F[2,j2]+1
    }
  }
}
F

n11=F[1,1]
n12=F[1,2]
n21=F[2,1]
n22=F[2,2]
n1=n11+n12
n2=n21+n22

P_esti=matrix(0,nrow=m,ncol=m) # Matriz de probabilidades de transicion
for (i in 1:m)
{
  for (j in 1:m)
  {
    P_esti[j,i]=F[j,i]/sum(F[j,])
  }
}
P_esti

# Grafica de R(P[1,1]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.70,0.90,by=0.001)
R=(p/P_esti[1,1])^(n11)*((1-p)/(1-P_esti[1,1]))^(n1-n11)

```

```

par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[11]),ylab=expression("R"*p[11]*";x"))
legend(0.72,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","
green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[1,1])

# Intervalos de confianza para P[1,1]

f1 <- function (x, a) (x/a)^(n11)*((1-x)/(1-a))^(n1-n11)-0.036
xmin1 <- uniroot(f1, c(0.70, P_esti[1,1]), tol = 0.0001, a=P_esti[1,1])
  $root
xmax1 <- uniroot(f1, lower=P_esti[1,1], upper=0.90, tol = 0.0001, a=P_esti
[1,1])$root
I1<-c(xmin1,xmax1)
I1

f2 <- function (x, a) (x/a)^(n11)*((1-x)/(1-a))^(n1-n11)-0.15
xmin2 <- uniroot(f2, c(0.70, P_esti[1,1]), tol = 0.0001, a=P_esti[1,1])
  $root
xmax2 <- uniroot(f2, lower=P_esti[1,1], upper=0.90, tol = 0.0001, a=P_esti
[1,1])$root
I2<-c(xmin2,xmax2)
I2

f3 <- function (x, a) (x/a)^(n11)*((1-x)/(1-a))^(n1-n11)-0.25
xmin3 <- uniroot(f3, c(0.70, P_esti[1,1]), tol = 0.0001, a=P_esti[1,1])
  $root
xmax3 <- uniroot(f3, lower=P_esti[1,1], upper=0.90, tol = 0.0001, a=P_esti
[1,1])$root
I3<-c(xmin3,xmax3)
I3

# Grafica de R(P[1,2]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.10,0.30,by=0.001)
R=(p/P_esti[1,2])^(n12)*((1-p)/(1-P_esti[1,2]))^(n1-n12)
par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[12]),ylab=expression("R"*p[12]*";x"))
legend(0.25,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","
green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")

```

```

abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[1,2])

# Intervalos de confianza para P[1,2]

f1 <- function (x, a) (x/a)^(n12)*((1-x)/(1-a))^(n1-n12)-0.036
xmin1 <- uniroot(f1, c(0.1, P_esti[1,2]), tol = 0.0001, a=P_esti[1,2])$root
xmax1 <- uniroot(f1, lower=P_esti[1,2], upper=0.30, tol = 0.0001, a=P_esti
  [1,2])$root
I1<-c(xmin1,xmax1)
I1

f2 <- function (x, a) (x/a)^(n12)*((1-x)/(1-a))^(n1-n12)-0.15
xmin2 <- uniroot(f2, c(0.1, P_esti[1,2]), tol = 0.0001, a=P_esti[1,2])$root
xmax2 <- uniroot(f2, lower=P_esti[1,2], upper=0.30, tol = 0.0001, a=P_esti
  [1,2])$root
I2<-c(xmin2,xmax2)
I2

f3 <- function (x, a) (x/a)^(n12)*((1-x)/(1-a))^(n1-n12)-0.25
xmin3 <- uniroot(f3, c(0.1, P_esti[1,2]), tol = 0.0001, a=P_esti[1,2])$root
xmax3 <- uniroot(f3, lower=P_esti[1,2], upper=0.30, tol = 0.0001, a=P_esti
  [1,2])$root
I3<-c(xmin3,xmax3)
I3

# Grafica de R(P[2,1]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.4,0.7,by=0.001)
R=(p/P_esti[2,1])^(n21)*((1-p)/(1-P_esti[2,1]))^(n2-n21)
par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[21]),ylab=expression("R"*p[21]*";x"))
legend(0.41,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","
  green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[2,1])

# Intervalos de confianza para P[2,1]

f1 <- function (x, a) (x/a)^(n21)*((1-x)/(1-a))^(n2-n21)-0.036

```

```

xmin1 <- uniroot(f1, c(0.4, P_esti[2,1]), tol = 0.0001, a=P_esti[2,1])$root
xmax1 <- uniroot(f1, lower=P_esti[2,1], upper=0.75, tol = 0.0001, a=P_esti
  [2,1])$root
I1<-c(xmin1,xmax1)
I1

f2 <- function (x, a) (x/a)^(n21)*((1-x)/(1-a))^(n2-n21)-0.15
xmin2 <- uniroot(f2, c(0.4, P_esti[2,1]), tol = 0.0001, a=P_esti[2,1])$root
xmax2 <- uniroot(f2, lower=P_esti[2,1], upper=0.75, tol = 0.0001, a=P_esti
  [2,1])$root
I2<-c(xmin2,xmax2)
I2

f3 <- function (x, a) (x/a)^(n21)*((1-x)/(1-a))^(n2-n21)-0.25
xmin3 <- uniroot(f3, c(0.4, P_esti[2,1]), tol = 0.0001, a=P_esti[2,1])$root
xmax3 <- uniroot(f3, lower=P_esti[2,1], upper=0.75, tol = 0.0001, a=P_esti
  [2,1])$root
I3<-c(xmin3,xmax3)
I3

# Grafica de R(P[2,2]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.25,0.65,by=0.001)
R=(p/P_esti[2,2])^(n22)*((1-p)/(1-P_esti[2,2]))^(n2-n22)
par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[22]),ylab=expression("R"*p[22]*";x"))
legend(0.51,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","
  green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[2,2])

# Intervalos de confianza para P[2,2]

f1 <- function (x, a) (x/a)^(n22)*((1-x)/(1-a))^(n2-n22)-0.036
xmin1 <- uniroot(f1, c(0.25, P_esti[2,2]), tol = 0.0001, a=P_esti[2,2])
  $root
xmax1 <- uniroot(f1, lower=P_esti[2,2], upper=0.6, tol = 0.0001, a=P_esti
  [2,2])$root
I1<-c(xmin1,xmax1)
I1

f2 <- function (x, a) (x/a)^(n22)*((1-x)/(1-a))^(n2-n22)-0.15

```

```

xmin2 <- uniroot(f2, c(0.25, P_esti[2,2]), tol = 0.0001, a=P_esti[2,2])
  $root
xmax2 <- uniroot(f2, lower=P_esti[2,2], upper=0.6, tol = 0.0001, a=P_esti
  [2,2])$root
I2<-c(xmin2,xmax2)
I2

f3 <- function (x, a) (x/a)^(n22)*((1-x)/(1-a))^(n2-n22)-0.25
xmin3 <- uniroot(f3, c(0.25, P_esti[2,2]), tol = 0.0001, a=P_esti[2,2])
  $root
xmax3 <- uniroot(f3, lower=P_esti[2,2], upper=0.6, tol = 0.0001, a=P_esti
  [2,2])$root
I3<-c(xmin3,xmax3)

```

8. Código en R para calcular la probabilidad de cobertura para el caso no paramétrico

```

S=5000
k=1
T=c(1,2)
T
m=length(T)
m
N=1826
cont11.1=0
cont11.2=0
cont11.3=0
cont12.1=0
cont12.2=0
cont12.3=0
cont21.1=0
cont21.2=0
cont21.3=0
cont22.1=0
cont22.2=0
cont22.3=0
xmin11.1=rep(0,N)
xmax11.1=rep(0,N)
xmin11.2=rep(0,N)
xmax11.2=rep(0,N)
xmin11.3=rep(0,N)
xmax11.3=rep(0,N)
xmin12.1=rep(0,N)
xmax12.1=rep(0,N)

```

```

xmin12.2=rep(0,N)
xmax12.2=rep(0,N)
xmin12.3=rep(0,N)
xmax12.3=rep(0,N)
xmin21.1=rep(0,N)
xmax21.1=rep(0,N)
xmin21.2=rep(0,N)
xmax21.2=rep(0,N)
xmin21.3=rep(0,N)
xmax21.3=rep(0,N)
xmin22.1=rep(0,N)
xmax22.1=rep(0,N)
xmin22.2=rep(0,N)
xmax22.2=rep(0,N)
xmin22.3=rep(0,N)
xmax22.3=rep(0,N)

while (k<S)
{
# Simulacion de una cadena de Markov en {1, 2} dadas una
# condicion inicial pi_0 y una matriz de transicion P

pi_0=c(0.762,0.238)
pi_0      # Distribucion inicial
P=matrix(c(0.817,0.585,0.183,0.415),nrow=2,ncol=2) # Matriz de transicion
P
N=1826    # Numero de pasos a simular (N=Total de datos+1)
X=matrix(0,nrow=1,ncol=N+1) # Inicializar y borrar si hay datos anteriores
t=0:N     # Indices de tiempo

# rando genera numeros aleatorios (1 o 2) con distribucion Uniforme([0,1])

rando<-function(p)
{
u=runif(1,min=0,max=1)
s=p[1]
j=1

while ((u>s) & (j<length(p)))
{
j=j+1
s=s+p[j]
}

index=j

```

```

}
X[1]=rando(pi_0) # Genera el valor de X(0) con distribucion pi_0

for (i in 1:N)
{
  X[i+1]=rando(P[X[i],])
}

n=length(X)
n

# Calculo de las frecuencias para dos estados

F=matrix(0,nrow=n,ncol=n) # Matriz de frecuencias
for (i in 1:(n-1))
{
  for (j1 in 1:m)
  {
    if (X[i]==1 & X[i+1]==j1)
    {
      F[1,j1]=F[1,j1]+1
    }
  }
  for (j2 in 1:m)
  {
    if (X[i]==2 & X[i+1]==j2)
    {
      F[2,j2]=F[2,j2]+1
    }
  }
}
F

n11=F[1,1]
n12=F[1,2]
n21=F[2,1]
n22=F[2,2]
n1=n11+n12
n2=n21+n22

P_esti=matrix(0,nrow=m,ncol=m) # Matriz de probabilidades de transicion
for (i in 1:m)
{
  for (j in 1:m)

```

```

    {
      P_esti[j,i]=F[j,i]/sum(F[j,])
    }
  }
P_esti

# Grafica de R(P[1,1]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.70,0.90,by=0.001)
R=(p/P_esti[1,1])^(n11)*((1-p)/(1-P_esti[1,1]))^(n1-n11)
par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[11]),ylab=expression("R"*p[11]*";x"))
legend(0.72,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[1,1])

# Intervalos de confianza para P[1,1]

f1 <- function (x, a) (x/a)^(n11)*((1-x)/(1-a))^(n1-n11)-0.036
xmin11.1[k] <- uniroot(f1, c(0.70, P_esti[1,1]), tol = 0.0001, a=P_esti
  [1,1])$root
xmax11.1[k] <- uniroot(f1, lower=P_esti[1,1], upper=0.90, tol = 0.0001, a=
  P_esti[1,1])$root

f2 <- function (x, a) (x/a)^(n11)*((1-x)/(1-a))^(n1-n11)-0.15
xmin11.2[k] <- uniroot(f2, c(0.70, P_esti[1,1]), tol = 0.0001, a=P_esti
  [1,1])$root
xmax11.2[k] <- uniroot(f2, lower=P_esti[1,1], upper=0.90, tol = 0.0001, a=
  P_esti[1,1])$root

f3 <- function (x, a) (x/a)^(n11)*((1-x)/(1-a))^(n1-n11)-0.25
xmin11.3[k] <- uniroot(f3, c(0.70, P_esti[1,1]), tol = 0.0001, a=P_esti
  [1,1])$root
xmax11.3[k] <- uniroot(f3, lower=P_esti[1,1], upper=0.90, tol = 0.0001, a=
  P_esti[1,1])$root

if (P[1,1]>xmin11.1[k] & P[1,1]<xmax11.1[k])
{
  cont11.1=cont11.1+1
}
if (P[1,1]>xmin11.2[k] & P[1,1]<xmax11.2[k])

```

```

{
  cont11.2=cont11.2+1
}
if (P[1,1]>xmin11.3[k] & P[1,1]<xmax11.3[k])
{
  cont11.3=cont11.3+1
}

# Grafica de R(P[1,2]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.10,0.30,by=0.001)
R=(p/P_esti[1,2])^(n12)*((1-p)/(1-P_esti[1,2]))^(n1-n12)
par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[12]),ylab=expression("R"*p[12]*";x"))
legend(0.25,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[1,2])

# Intervalos de confianza para P[1,2]

f1 <- function (x, a) (x/a)^(n12)*((1-x)/(1-a))^(n1-n12)-0.036
xmin12.1[k] <- uniroot(f1, c(0.1, P_esti[1,2]), tol = 0.0001, a=P_esti
  [1,2])$root
xmax12.1[k] <- uniroot(f1, lower=P_esti[1,2], upper=0.30, tol = 0.0001, a=
  P_esti[1,2])$root

f2 <- function (x, a) (x/a)^(n12)*((1-x)/(1-a))^(n1-n12)-0.15
xmin12.2[k] <- uniroot(f2, c(0.1, P_esti[1,2]), tol = 0.0001, a=P_esti
  [1,2])$root
xmax12.2[k] <- uniroot(f2, lower=P_esti[1,2], upper=0.30, tol = 0.0001, a=
  P_esti[1,2])$root

f3 <- function (x, a) (x/a)^(n12)*((1-x)/(1-a))^(n1-n12)-0.25
xmin12.3[k] <- uniroot(f3, c(0.1, P_esti[1,2]), tol = 0.0001, a=P_esti
  [1,2])$root
xmax12.3[k] <- uniroot(f3, lower=P_esti[1,2], upper=0.30, tol = 0.0001, a=
  P_esti[1,2])$root

if (P[1,2]>xmin12.1[k] & P[1,2]<xmax12.1[k])
{
  cont12.1=cont12.1+1
}

```

```

}
if (P[1,2] > xmin12.2[k] & P[1,2] < xmax12.2[k])
{
  cont12.2=cont12.2+1
}
if (P[1,2] > xmin12.3[k] & P[1,2] < xmax12.3[k])
{
  cont12.3=cont12.3+1
}

# Grafica de R(P[2,1]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.4,0.7,by=0.001)
R=(p/P_esti[2,1])^(n21)*((1-p)/(1-P_esti[2,1]))^(n2-n21)
par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[21]),ylab=expression("R"*p[21]*";x"))
legend(0.41,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[2,1])

# Intervalos de confianza para P[2,1]

f1 <- function(x, a) (x/a)^(n21)*((1-x)/(1-a))^(n2-n21)-0.036
xmin21.1[k] <- uniroot(f1, c(0.4, P_esti[2,1]), tol = 0.0001, a=P_esti[2,1])$root
xmax21.1[k] <- uniroot(f1, lower=P_esti[2,1], upper=0.75, tol = 0.0001, a=P_esti[2,1])$root

f2 <- function(x, a) (x/a)^(n21)*((1-x)/(1-a))^(n2-n21)-0.15
xmin21.2[k] <- uniroot(f2, c(0.4, P_esti[2,1]), tol = 0.0001, a=P_esti[2,1])$root
xmax21.2[k] <- uniroot(f2, lower=P_esti[2,1], upper=0.75, tol = 0.0001, a=P_esti[2,1])$root

f3 <- function(x, a) (x/a)^(n21)*((1-x)/(1-a))^(n2-n21)-0.25
xmin21.3[k] <- uniroot(f3, c(0.4, P_esti[2,1]), tol = 0.0001, a=P_esti[2,1])$root
xmax21.3[k] <- uniroot(f3, lower=P_esti[2,1], upper=0.75, tol = 0.0001, a=P_esti[2,1])$root

if (P[2,1] > xmin21.1[k] & P[2,1] < xmax21.1[k])

```

```

{
  cont21.1=cont21.1+1
}
if (P[2,1]>xmin21.2[k] & P[2,1]<xmax21.2[k])
{
  cont21.2=cont21.2+1
}
if (P[2,1]>xmin21.3[k] & P[2,1]<xmax21.3[k])
{
  cont21.3=cont21.3+1
}

# Grafica de R(P[2,2]) y los niveles para c=0.25, 0.15 y 0.036

p=seq(0.25,0.65,by=0.001)
R=(p/P_esti[2,2])^(n22)*((1-p)/(1-P_esti[2,2]))^(n2-n22)
par(mfrow=c(1,1))
plot(p,R,type="l",xlab=expression(p[22]),ylab=expression("R"*p[22]*";x"))
legend(0.51,0.8,c("c=0.25","c=0.15","c=0.036"),cex=0.8,col=c("red","blue","green"),lty=1)
title(main="FUNCION DE VEROSIMILITUD RELATIVA")
abline(a=0.25,b=0,col="red")
abline(a=0.15,b=0,col="blue")
abline(a=0.036,b=0,col="green")
abline(v=P_esti[2,2])

# Intervalos de confianza para P[2,2]

f1 <- function(x, a) (x/a)^(n22)*((1-x)/(1-a))^(n2-n22)-0.036
xmin22.1[k] <- uniroot(f1, c(0.25, P_esti[2,2]), tol = 0.0001, a=P_esti[2,2])$root
xmax22.1[k] <- uniroot(f1, lower=P_esti[2,2], upper=0.6, tol = 0.0001, a=P_esti[2,2])$root

f2 <- function(x, a) (x/a)^(n22)*((1-x)/(1-a))^(n2-n22)-0.15
xmin22.2[k] <- uniroot(f2, c(0.25, P_esti[2,2]), tol = 0.0001, a=P_esti[2,2])$root
xmax22.2[k] <- uniroot(f2, lower=P_esti[2,2], upper=0.6, tol = 0.0001, a=P_esti[2,2])$root

f3 <- function(x, a) (x/a)^(n22)*((1-x)/(1-a))^(n2-n22)-0.25
xmin22.3[k] <- uniroot(f3, c(0.25, P_esti[2,2]), tol = 0.0001, a=P_esti[2,2])$root
xmax22.3[k] <- uniroot(f3, lower=P_esti[2,2], upper=0.6, tol = 0.0001, a=P_esti[2,2])$root

```

```

if (P[2,2] > xmin22.1[k] & P[2,2] < xmax22.1[k])
{
    cont22.1=cont22.1+1
}
if (P[2,2] > xmin22.2[k] & P[2,2] < xmax22.2[k])
{
    cont22.2=cont22.2+1
}
if (P[2,2] > xmin22.3[k] & P[2,2] < xmax22.3[k])
{
    cont22.3=cont22.3+1
}

k=k+1
}

# Probabilidades de cobertura a los niveles c=0.036, 0.15 y 0.25

PC11.1=cont11.1/S
PC11.1
PC11.2=cont11.2/S
PC11.2
PC11.3=cont11.3/S
PC11.3
PC12.1=cont12.1/S
PC12.1
PC12.2=cont12.2/S
PC12.2
PC12.3=cont12.3/S
PC12.3
PC21.1=cont21.1/S
PC21.1
PC21.2=cont21.2/S
PC21.2
PC21.3=cont21.3/S
PC21.3
PC22.1=cont22.1/S
PC22.1
PC22.2=cont22.2/S
PC22.2
PC22.3=cont22.3/S
PC22.3

```

Bibliografía

- [1] Anderson, T. y Goodman, L. (1957). Statistical Inference about Markov Chains. *The Annals of Mathematical Statistics*. págs. 89-110.
- [2] Barnard, G. y Sprott, D. (2006). Likelihood. *Encyclopedia of Statistical Science* (Second edition), págs 4165-4170. Editado por Kotz, S. John Wiley & Sons, Inc.
- [3] Barrantes, J. y Vega, N. (2002). La persistencia y las cadenas de Markov aplicadas a la lluvia en San José durante el periodo 1950-1990. *Tópicos Meteorológicos y Oceanográficos*, págs. 12-19.
- [4] Basawa, I. y Prakasa, B. L. S. (1980). Asymptotic Inference for Stochastic Processes. *Stochastic Processes and their Applications*, págs. 221-254. North-Holland Publishing Company.
- [5] Cazacioc, L. y Corina, E. (2005). Evaluation of the transition probabilities for daily precipitation time series using a Markov chain model. *The 3-rd International Colloquium "Mathematics in Engineering and Numerical Physics"*, págs. 82-92.
- [6] Craig, B. y Sendi, P. (2002). Estimation of the Transition Matrix of a Discrete-time Markov Chain. *Health Economics*, págs. 33-42.
- [7] Edwards, A. (2006). *Likelihood* (Extended edition). The Johns Hopkins University Press.
- [8] Evans, M. y Roshental, J. (2005). *Probabilidad y estadística*. Editorial Reverté S. A.
- [9] Faría, A. y Puche, M. (1997). Estudio de periodos secos del Municipio de La Cadaña de Urdaneta del Estado de Zulia, Venezuela. *Revista Facultad de Agronomía UCV*, págs. 21-31.
- [10] Fisher, R. (1921). On the "Probable Error" of a Coefficient of Correlation from a Small Sample. *Metron*, V. 1, págs. 3-32.

- [11] Gregory, J., Wigley, T. y Jones, P. (1992). Determining and Interpreting the Order of a Two-State Markov Chain: Application to Models of Daily Precipitation. *Water Resources Research*, págs. 1443-1446.
- [12] Haaser, N., LaSalle, J. y Sullivan, J. (2012). *Análisis matemático*. V. 1. Editorial Trillas.
- [13] Hoel, P., Port, S. y Stone, C. (1972). *Introduction to Stochastic Processes*. Houghton Mifflin Company.
- [14] Inga, R. (1996). Principio de maximización de la información de Akaike. *Revista de Ciencias - UNI*, págs. 31-40.
- [15] Kalbfleisch, J. (1979). *Probability and Statistical Inference*. V. 2. Springer-Verlag.
- [16] Karlin, S. y Taylor, H. (1975). *A First Course in Stochastic Processes* (Second edition). Academic Press, Inc.
- [17] Lawless, J. (2003). *Statistical Models and Methods for Lifetime Data* (Second edition). John Wiley & Sons, Inc.
- [18] Liao, Y., Zhang, Q. y Chen, D. (2004). Stochastic modeling of daily precipitation in China. *Journal of Geographical Sciences*, págs. 417-426.
- [19] Lindsey, J. (1999). Some Statistical Heresies. *The Statistician* V. 48, págs. 1-40.
- [20] Luenberger, D. (1989). *Programación lineal y no lineal*. Addison-Wesley Iberoamericana.
- [21] Luengo, M., Ceballos, A., Martínez, J. y Yuste, C. (2002). Las rachas secas en el sector central de la cuenca del Duero. *Investigaciones geográficas*, págs. 65-82.
- [22] Martínez, A., Burton, A., Sordo, A. y Garrote, L. (2004). Evaluación de modelos multivariados para la simulación estocástica de lluvias a escala diaria. *Revista Ingeniería del Agua*.
- [23] Martínez, J. y Villalón, J. (2003). *Introducción al cálculo estocástico aplicado a la modelización económico-financiero-actuarial*. Editorial Netbiblo.
- [24] Meeker, W. y Escobar, L. (1998). *Statistical Methods for Reliability Data*. John Wiley & Sons.
- [25] Menéndez, F y Lara, C. (2011). *Teoría de integración*. Serie: Textos. V. 13. Publicaciones Electrónicas. Sociedad Matemática Mexicana.
- [26] Mood, A. y Graybill, F. (1978). *Introducción a la teoría de la estadística*. Ediciones Aguilar S. A.

- [27] Rincón, L. (2011). *Introducción a los procesos estocásticos*. Departamento de Matemáticas. Facultad de Ciencias UNAM.
- [28] Roshental, J. (2006). *At Rigorous Theory Probability* (Second edition). World Scientific Publishing Co.
- [29] Rosselló, V. et al. (2004). *Historia, clima y paisaje. Estudios geográficos en memoria del profesor Antonio López Gómez*. Universidad de Valencia. Artes Gráficas Soler, S. L.
- [30] Salazar, J. et al. (2007). Modelo de Markov de tres estados: comparación de parametrizaciones de la tasa de intensidad de transición. Aplicación a datos de artritis reumatoidea. *Revista Colombiana de Estadística*, págs. 213-299.
- [31] Serfling, D. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons, Inc.
- [32] Skuriat-Olechnowska, M. (2005). *Statistical Inference and Hypothesis Testing for Markov Chains with Interval Censoring. Application to bridge condition data in the Netherlands*. Thesis of Master Science, Delft University of Tecnology, Delft, Netherlands.
- [33] Sprott, D. (2000). *Statistical Inference in Science*. Springer-Verlag New York, Inc.
- [34] Tovar, R. y Salazar, J. (2009). Un modelo lineal mixto adaptado a una cadena de Markov con espacio de estados ordinal. Aplicación a datos sobre promedios académicos de estudiantes. *Revista Colombiana de Estadística*, págs. 213-230.
- [35] Zavaleta, J. (2007). *Aplicaciones Computacionales de las EDP's Asociadas a Procesos Estocásticos*. Tesis de Licenciatura en Matemáticas Aplicadas, UTM, Huajuapán de León, Oaxaca, México.