

# Aspectos básicos en la Inferencia Estadística para Cadenas de Markov en tiempo discreto

José Batún-Cutz<sup>1</sup>, Ruben Cool-Padilla<sup>2</sup>, Ernesto Guerrero-Lara<sup>3</sup>, Diódora Kantún-Chim<sup>4</sup> y Henry Pantí-Trejo<sup>5</sup>

Facultad de Matemáticas, Universidad Autónoma de Yucatán.

<sup>1</sup>jbatun@correo.uady.mx, <sup>2</sup>ruben.cool@correo.uady.mx, <sup>3</sup>ernesto.guerrero@correo.uady.mx,

<sup>4</sup>kchim@correo.uady.mx, <sup>5</sup>henry.panti@correo.uady.mx

## Resumen

*El presente artículo tiene como objetivo presentar un panorama general sobre la inferencia estadística para las cadenas de Markov en tiempo discreto, utilizando artículos y libros publicados sobre estos modelos. Los temas que en este artículo se abordan incluyen: estimación puntual y por intervalos para las probabilidades de transición y una prueba de hipótesis para el supuesto de homogeneidad en una cadena de Markov. A lo largo del artículo se presentan ejemplos tomados de la literatura, en los que se aplica el modelo de cadenas de Markov para resolver problemas de interés, y que sirven de ilustración para propósitos de este trabajo. Este trabajo espera brindar apoyo a personas interesadas en un primer contacto con las cadenas de Markov y que sirva de guía en su búsqueda de información sobre el tema.*

**Palabras Clave:** Cadenas de Markov, estados absorbentes, estimación por máxima verosimilitud, intervalos de confianza, razón de verosimilitud.

DOI: 10.36788/sah.v6i1.131

Recibido: 14 de julio de 2022

Aceptado: 5 de septiembre de 2022

## Introducción

Ha transcurrido más de un siglo desde que el ilustre matemático ruso Andréi Markov sentó las bases de la teoría que permitió el estudio de los objetos matemáticos que el día de hoy llevan su nombre. La teoría desarrollada para las cadenas de Markov ha sido de gran utilidad en el avance de la Probabilidad hasta nuestros días, principalmente en el estudio de los procesos estocásticos, y es que las cadenas de Markov son consideradas entre los principales modelos que aluden cuando se trata de procesos estocásticos.

Desde el punto de vista teórico, las cadenas de Markov contemplan cálculos matemáticos sencillos para la obtención de fórmulas para cantidades que son de interés en el fenómeno de estudio. Al igual, desde el punto de vista práctico, las cadenas de Markov han mostrado ser modelos adecuados para fenómenos aleatorios que surgen en diferentes áreas del conocimiento, por ejemplo, en [20] utilizan cadenas de Markov para modelar el surgimiento de discusiones en la red social de Twitter y que se relacionan con la enfermedad del Coronavirus. Por otra

parte, en [13] se modela la dinámica de células cancerosas en cada ciclo proliferativo mediante una cadena de Markov y en [16] modelan la retinopatía diabética con una cadena de Markov con cinco estados ordenados de acuerdo a la gravedad de la enfermedad. En genética, [2] citado en [9] presentan un estudio en el que se describe la dinámica de las bases nitrogenadas del ADN de la primera secuencia intermedia del gen del preproglucagón humano mediante cadenas de Markov.

La versatilidad del modelo se debe a su principal característica de no guardar memoria del pasado, es decir, la probabilidad de algún comportamiento futuro depende únicamente de la información que se tenga en el presente y no es de importancia la información del pasado. Esta propiedad, permite que las cadenas de Markov sean un modelo inmediato de asociar cuando se desea estudiar un fenómeno que cumpla con esta característica. Además, desde el punto de vista estadístico la aplicación de estos modelos mayormente son fáciles de implementar y los cálculos no requieren herramientas computacionales avanzadas.

Desde el punto de vista estadístico, uno de los primeros tratamientos de la teoría inferencial para cadenas de Markov que se conoce fue planteado en [1]. En ese trabajo se desarrolla de manera general la teoría de estimación puntual y pruebas de hipótesis para cadenas de Markov homogéneas y parcialmente las no homogéneas. Los resultados en ese artículo fueron el punto de partida para el avance de la teoría inferencial, utilizando herramientas más avanzadas o de actualidad como lo es el bootstrap.

El principal propósito de este artículo es presentar la teoría inferencial básica desarrollada para las cadenas de Markov de una manera, que en la opinión de los autores, es sencilla y accesible. Se presenta la estimación puntual, vía máxima verosimilitud, estimación por intervalos para los parámetros que definen a una cadena de Markov. De igual forma, se presenta una prueba de hipótesis para el supuesto de homogeneidad de una cadena de Markov, presente en la mayoría de las implementaciones de este modelo. Todo lo anterior, es ilustrado mediante ejemplos tomados de la literatura.

Para alcanzar el objetivo, el presente artículo está organizado en tres secciones. En la Sección 1 se presentan los preliminares, es decir, la teoría central de las cadenas de Markov, algunos teoremas y proposiciones que son de utilidad para la construcción de la teoría inferencial en la modelación ([18]). La Sección 2 se divide en dos: Estimación y Prueba de Hipótesis. En la primera parte se presentan la función de verosimilitud y los estimadores de máxima verosimilitud para los parámetros del modelo de cadenas de Markov homogéneas ([1], [3], [9]). En la segunda parte se muestra una prueba de hipótesis para la homogeneidad de la cadena de Markov; esto se realiza considerando también las cadenas de Markov no homogéneas planteando su función de máxima verosimilitud y los estimadores de máxima verosimilitud para este tipo de cadenas ([1]). Finalmente, en la Sección 3 se presentan algunas conclusiones y comentarios del artículo.

## 1. Cadenas de Markov

Un **proceso estocástico en tiempo discreto** es una colección de variables aleatorias  $\{X_t, t \in T\}$  definidas sobre un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ , indexada por un conjunto

contable  $T$ , llamado *tiempo* y los valores que toma cada variable aleatoria  $X_t$  conforman al conjunto  $S$  llamado *espacio de estados*. Es usual considerar el conjunto  $T = \{0, 1, 2, 3, \dots\}$  y escribir  $\{X_n, n \geq 0\}$  para denotar al proceso estocástico en tiempo discreto. El espacio de estados  $S$  puede ser contable o no contable, sin embargo, en este trabajo supondremos que el conjunto  $S$  es finito y escribiremos  $S = \{1, \dots, s\}$ .

Las *trayectorias* o *realizaciones* de un proceso estocástico son útiles para visualizar gráficamente su comportamiento a través del tiempo. Una trayectoria es una función  $n \mapsto X_n(\omega)$ , donde  $\omega \in \Omega$  es fijo. De esta forma, para cada  $\omega \in \Omega$  tenemos una trayectoria de  $\{X_n, n \geq 0\}$ . En la Figura 1 se presentan las gráficas de dos trayectorias de un proceso estocástico con espacio de estados  $S = \{1, 2, 3, 4, 5\}$ .

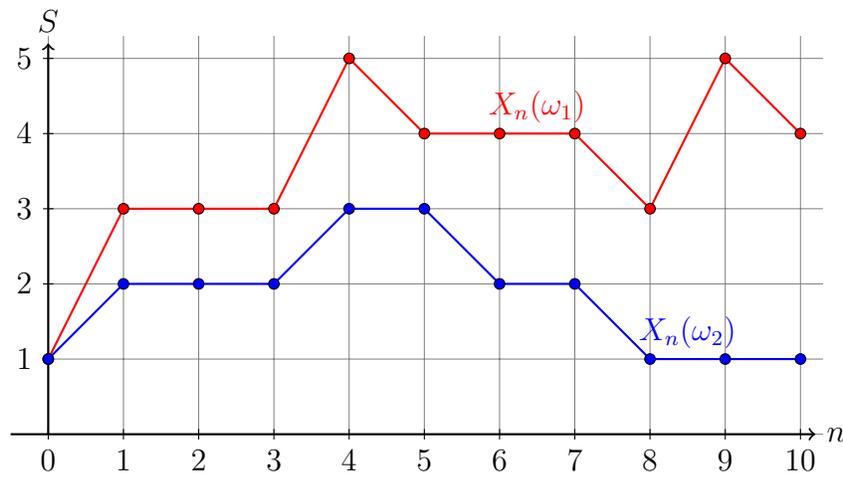


Figura 1: Ejemplo de trayectorias de un proceso estocástico  $\{X_n, n \geq 0\}$

Ya que los procesos estocásticos están definidos por variables aleatorias, éstos pueden ser considerados como cantidades que varían aleatoriamente a través del tiempo y con base en el comportamiento probabilístico de las variables aleatorias que los definen, se pueden catalogar en diferentes tipos de procesos. Son de interés en este artículo los procesos estocásticos en los que la probabilidad de cualquier comportamiento inmediato futuro del proceso solo depende del estado actual y no de su pasado. Este tipo de procesos son conocidos como cadenas de Markov y como se ilustrará más adelante, tienen una amplia aplicación en la modelación de diversos fenómenos aleatorios. De manera formal, definimos una cadena de Markov:

**Definición 1.** Una *cadena de Markov* es un proceso estocástico  $\{X_n : n \geq 0\}$  con espacio de estados  $S = \{1, \dots, s\}$  que satisface la **propiedad de Markov**, esto es, para cualquier entero  $n \geq 0$  y para cualesquiera estados  $s_0, \dots, s_{n-1}, i, j$ , se cumple:

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = s_{n-1}, \dots, X_0 = s_0) = P(X_{n+1} = j \mid X_n = i). \quad (1)$$

La propiedad de Markov establece una propiedad de pérdida de memoria debido a que no importa los estados previos donde la cadena de Markov estuvo sino solo se requiere el

conocimiento de la posición del instante actual para conocer la probabilidad de que la cadena cambie de estado o permanezca, en el instante siguiente.

Si el lado derecho de la ecuación (1) no depende de  $n$  se dice que la cadena de Markov es *homogénea* en el tiempo, en cuyo caso se escribe  $p_{ij} = P(X_{n+1} = j \mid X_n = i)$ . La probabilidad  $p_{ij}$  es la probabilidad de que la cadena de Markov transite del estado  $i$  al estado  $j$  en una unidad de tiempo, esta probabilidad es conocida como *probabilidad de transición*. Las probabilidades de transición se pueden organizar en una matriz cuadrada  $\mathbf{P} = (p_{ij})$  llamada *matriz de transición*. Los elementos de esta matriz satisfacen las siguientes dos propiedades:

1.  $0 \leq p_{ij} \leq 1$ , para todo  $i, j \in S$ .
2.  $\sum_{j \in S} p_{ij} = 1$ , para toda  $i \in S$ .

Recíprocamente, toda matriz que cumpla las Propiedades 1 y 2 es llamada matriz de transición.

La **distribución inicial** de una cadena de Markov es la distribución de probabilidad de la variable aleatoria  $X_0$ , esto es, considerando  $p_i = P(X_0 = i)$ , para  $i \in S$ , el vector de probabilidades  $\mathbf{p}_0 = (p_1, p_2, \dots, p_s)$  corresponde a la distribución inicial de la cadena de Markov. Esta distribución proporciona información del punto de inicio o partida de la cadena de Markov.

Las probabilidades de transición  $p_{ij}$  y los elementos de la distribución inicial  $p_1, p_2, \dots, p_s$  son los parámetros del modelo de cadenas de Markov. En algunas aplicaciones se conoce el punto inicial de las cadenas de Markov, lo que implica que solo se requiere estimar los elementos de la matriz de transición  $\mathbf{P}$  y en este caso, con base en la Propiedad 2, se estiman  $s(s-1)$  parámetros desconocidos.

La relevancia de la matriz de transición y la distribución inicial en una cadena de Markov radica en que ambas determinan completamente el comportamiento probabilístico de la cadena de Markov. Este resultado, establecido en la siguiente proposición, se puede encontrar en [18].

**Proposición 1.** *Sea  $\{X_n, n \geq 0\}$  una cadena de Markov con espacio de estados  $S$ . Para cualesquiera  $n \geq 0$  y  $s_0, s_1, \dots, s_n \in S$ , se satisface:*

$$P(X_0 = s_0, X_1 = s_1, \dots, X_{n-1} = s_{n-1}, X_n = s_n) = p_{s_0} \cdot p_{s_0 s_1} \cdots p_{s_{n-2}, s_{n-1}} p_{s_{n-1}, s_n}.$$

La proposición anterior también establece la probabilidad de observar una determinada trayectoria en términos de los elementos de la matriz de transición  $\mathbf{P}$  y la distribución inicial  $\mathbf{p}_0$  de la cadena de Markov, lo cual a su vez es de utilidad para establecer la función de verosimilitud de los parámetros del modelo de cadenas de Markov.

Durante el modelado de un fenómeno, una posible cantidad de interés en las cadenas de Markov homogéneas es la probabilidad de que, estando en el estado  $i$ , después de haber transcurrido  $n$  unidades de tiempo la cadena de Markov se encuentre en el estado  $j$ , es decir,  $P(X_{m+n} = j \mid X_m = i)$ , para  $m \in \{0, 1, 2, \dots\}$ . A esta probabilidad se le denomina *probabilidad de transición en  $n$  pasos* y la denotamos como  $p_{ij}^{(n)}$ . De la definición se sigue que  $p_{ij}^{(1)} = p_{ij}$  y

si  $n = 0$  entonces  $p_{ij}^{(0)} = 1$  si  $i = j$  y  $p_{ij}^{(0)} = 0$  si  $i \neq j$ . Las probabilidades de transición en  $n$  pasos satisfacen las llamadas ecuaciones de Chapman-Kolmogorov:

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

para toda  $i, j \in S$ ,  $0 \leq r \leq n$ . Las probabilidades de transición en  $n$  pasos se pueden organizar en una matriz cuadrada  $\mathbf{P}^{(n)}$  conocida como *matriz de transición en  $n$  pasos* y con base en las ecuaciones de Chapman-Kolmogorov, se verifica que sus elementos pueden obtenerse a partir de los elementos de la matriz  $\mathbf{P}$ , esto se establece en la siguiente proposición, la cual se puede encontrar en [18].

**Proposición 2.** *La matriz de transición en  $n$  pasos, en una cadena de Markov homogénea, satisface  $\mathbf{P}^{(n)} = \mathbf{P}^n$ .*

La Proposición 2 establece que la probabilidad  $P(X_{m+n} = j | X_m = i)$ , para  $m \in \{0, 1, 2, \dots\}$  es el  $(i, j)$ -ésimo elemento de la  $n$ -ésima potencia de la matriz  $\mathbf{P}$ .

En el siguiente ejemplo se presenta una investigación en la cual se plantea un modelo de cadenas de Markov que describe la progresión mensual del VIH en un conjunto de sujetos que son susceptibles al riesgo de desarrollar una infección llamada Mycobacterium avium complex. Esta progresión se monitorea mensualmente y los estados que se consideran son tres rangos distintos de conteos de linfocitos CD4: 0-49, 50-74 y de 75 en adelante. Para este estudio la Proposición 2 permite calcular la probabilidad de que el paciente mantenga, aumente o disminuya su conteo de linfocitos CD4, después de haber transcurrido cierto número de meses.

**Ejemplo 1.** *En [8] se define la cadena de Markov  $\{X_n, n \geq 0\}$ , donde la variable aleatoria  $X_n$  representa el estado del paciente al mes  $n$  con base en el conteo de linfocitos CD4, es decir, el estado 1 es el rango 0-49, el estado 2 el rango 50-74 y el estado 3 el rango de 75 en adelante. En el estudio se está interesado en conocer la probabilidad de que un paciente al sexto mes tenga un conteo de linfocitos CD4 entre 0-49 dado que al iniciar el estudio su conteo de linfocitos CD4 se encontraba en el mismo rango, es decir, se está interesado en calcular  $p_{11}^{(6)} = P(X_6 = 1 | X_0 = 1)$ . Para calcular la probabilidad de interés es necesario considerar la Proposición 2, por lo que  $\mathbf{P}^{(6)}$  es calculada mediante la sexta potencia de la matriz  $\mathbf{P}$ . A continuación se muestran las matrices  $\mathbf{P}$  y  $\mathbf{P}^{(6)}$  del estudio:*

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 0.9630 & 0.0211 & 0.0159 \\ 0.6367 & 0.2123 & 0.1510 \\ 0.2119 & 0.2119 & 0.5762 \end{pmatrix} \end{matrix}, \quad \mathbf{P}^{(6)} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 0.9193 & 0.0360 & 0.0447 \\ 0.8966 & 0.0430 & 0.0604 \\ 0.8471 & 0.0582 & 0.0947 \end{pmatrix} \end{matrix}.$$

Con base en la matriz  $\mathbf{P}^{(6)}$ , se obtiene que  $p_{11}^{(6)} = 0.9193$ , es decir, la probabilidad de que al transcurrir 6 meses de estudio un paciente con un conteo de linfocitos entre 0-49 se encuentre en ese mismo rango es de 0.9193.

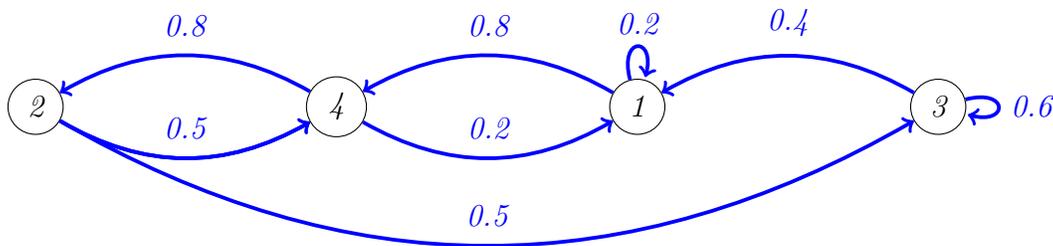
Una herramienta gráfica que permite visualizar de una forma más clara las transiciones entre estados en una cadena de Markov son los diagramas de transición, la cual es semejante a

un grafo dirigido. Un *diagrama de transición* se construye colocando los estados sin un orden en específico y en aquellos donde se tenga probabilidad de transición positiva, se dibuja una flecha orientada hacia la transición que se realiza, adicionalmente se escribe la probabilidad de transición sobre la flecha. Para ilustrar el diagrama de transición se presenta el siguiente ejemplo.

**Ejemplo 2.** Sea  $\{X_n, n \geq 0\}$  una cadena de Markov con matriz de transición dada por:

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 0.2 & 0 & 0 & 0.8 \\ 0 & 0 & 0.5 & 0.5 \\ 0.4 & 0 & 0.6 & 0 \\ 0.2 & 0.8 & 0 & 0 \end{pmatrix} \end{matrix}.$$

El diagrama de transición para esta cadena de Markov está dado por:



Se dice que la cadena de Markov es *irreducible* si desde cualquier estado se puede transitar a otro en un número finito de pasos. Los diagramas de transición permiten visualizar si una cadena de Markov es irreducible. En el diagrama del Ejemplo 2 se puede observar que del estado 1 se puede transitar al estado 4 en un paso con una probabilidad de 0.8. Del estado 3 se puede transitar al estado 2 en tres pasos, primero transitando al estado 1 con una probabilidad de 0.4, del estado 1 se transita al estado 4 con una probabilidad de 0.8 y por último, del estado 4 se transita al estado 2 con una probabilidad de 0.8. Un análisis similar se puede realizar para los estados 2 y 4. Lo anterior verifica que la cadena de Markov en el Ejemplo 2 es irreducible.

Un tipo de cadenas de Markov de especial interés son las llamadas cadenas de Markov con estados absorbentes. Un estado  $i$  es llamado absorbente si  $p_{ii} = 1$ , esto significa que si la cadena de Markov se encuentra o alcanza el estado  $i$ , se mantiene en este mismo estado en los tiempos siguientes. Si este es el caso, se dice que la cadena de Markov es absorbida en el estado  $i$ .

Existe una diversidad de estudios que utilizan las cadenas de Markov con estados absorbentes. En el ámbito educativo presentamos el ejemplo siguiente.

**Ejemplo 3.** En [6] se propone una cadena de Markov para modelar la trayectoria académica de estudiantes en una universidad en México. El espacio de estados está conformado por diez estados, los primeros ocho estados corresponden a los ocho semestres que conforman

la carrera, el estado 9 representa el estado de baja y el estado 10 representa el egreso de la carrera. La asignación del semestre se realiza con base en los créditos acumulados por el estudiante al paso del tiempo. A continuación se presenta la matriz de probabilidades de transición del modelo de cadenas de Markov de este estudio.

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{matrix} & \left( \begin{array}{cccccccccc} 0.390 & 0.431 & 0 & 0 & 0 & 0 & 0 & 0 & 0.179 & 0 \\ 0 & 0.371 & 0.482 & 0.008 & 0 & 0 & 0 & 0 & 0.139 & 0 \\ 0 & 0 & 0.377 & 0.532 & 0.028 & 0 & 0 & 0 & 0.063 & 0 \\ 0 & 0 & 0 & 0.335 & 0.618 & 0.004 & 0 & 0 & 0.043 & 0 \\ 0 & 0 & 0 & 0 & 0.355 & 0.551 & 0.030 & 0 & 0.064 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.303 & 0.632 & 0.027 & 0.038 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.214 & 0.648 & 0.038 & 0.101 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.285 & 0.020 & 0.695 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \end{matrix}.$$

Se puede observar en la matriz  $\mathbf{P}$  que las probabilidades  $p_{99}$  y  $p_{1010}$  son iguales a uno, esto quiere decir que los estados 9 y 10 son estados absorbentes porque si un estudiante causa baja del programa educativo se mantiene en ese estado en los tiempos siguientes, análogamente para el estado de egreso.

## 2. Inferencia Estadística

### 2.1 Estimación

El papel que juega la función de verosimilitud de una cadena de Markov es imprescindible para la estimación y pruebas de hipótesis en la teoría inferencial para este tipo de procesos estocásticos. La función de verosimilitud permite obtener los estimadores de máxima verosimilitud (EMV) para los elementos de la matriz de transición  $\mathbf{P}$  y mediante la razón de verosimilitud se puede establecer una prueba de hipótesis relativa al supuesto de homogeneidad de la cadena de Markov.

#### 2.1.1 Estimación puntual

Para obtener la función de verosimilitud es necesario entender cómo se presentan los datos, es decir, es fundamental definir una muestra aleatoria para una cadena de Markov. En [4] se sugiere la siguiente manera de visualizar una muestra aleatoria de una cadena de Markov. Sea  $h$  un número entero positivo. Suponga que  $\mathbf{x}$  es un conjunto finito de puntos obtenido de observar la cadena de Markov  $\{X_n, n \geq 0\}$ , desde el instante 0 hasta el tiempo  $h$ , es decir,  $\mathbf{x} = \{(0, s_0), (1, s_1), \dots, (h, s_h)\}$ , donde el primer elemento de la pareja  $(t, s_t)$  representa el tiempo y  $s_t$  es el estado en el que se encuentra el proceso en el tiempo  $t$ , con  $t = 0, 1, 2, \dots, h$ . En este caso,  $\mathbf{x}$  es llamado una trayectoria de longitud  $h$ . De esta manera, una muestra aleatoria de tamaño  $m$  de  $\{X_n, n \geq 0\}$  es una colección finita de  $m$

trayectorias independientes, de longitud  $h$  cada una, es decir, la muestra aleatoria está dada por  $\{\mathbf{x}_1, \dots, \mathbf{x}_m\}$ , donde:

$$\mathbf{x}_a = \{(0, s_0^{(a)}), (1, s_1^{(a)}), \dots, (h, s_h^{(a)})\}, \quad a = 1, \dots, m. \quad (2)$$

Nótese que en el Ejemplo 3, cada muestra inicia en el mismo estado, es decir,  $s_0^{(1)} = s_0^{(2)} = \dots = s_0^{(m)}$ . Es importante mencionar, que aun bajo el supuesto de homogeneidad, es posible obtener observaciones iniciales diferentes para cada trayectoria, pero esto no origina estimadores de máxima verosimilitud diferentes, ya que no importan los instantes en el tiempo en el que se hicieron las transiciones, sino el número de transiciones realizadas, como se mostrará a continuación.

Para establecer la función de verosimilitud, introducimos la siguiente notación. Sea  $n_{ij}^{(a)}$  el número de veces que se observó la transición del estado  $i$  al estado  $j$  en la  $a$ -ésima trayectoria de longitud  $h$  y sea  $n_{ij}$  el número de transiciones que se observan del estado  $i$  al estado  $j$  en la muestra. Entonces se satisface  $n_{ij} = \sum_{a=1}^m n_{ij}^{(a)}$ .

A manera de ilustración, consideremos una muestra aleatoria de una cadena de Markov con espacio de estados  $S = \{1, 2, 3\}$ , la cual está conformada por cuatro trayectorias de longitud diez:

$$\begin{aligned} \mathbf{x}_1 &= \{(0, 1), (1, 1), (2, 1), (3, 1), (4, 2), (5, 3), (6, 3), (7, 3), (8, 3), (9, 3), (10, 3)\} \\ \mathbf{x}_2 &= \{(0, 1), (1, 1), (2, 2), (3, 2), (4, 1), (5, 1), (6, 1), (7, 2), (8, 2), (9, 1), (10, 2)\} \\ \mathbf{x}_3 &= \{(0, 1), (1, 1), (2, 2), (3, 2), (4, 2), (5, 1), (6, 2), (7, 3), (8, 3), (9, 3), (10, 3)\} \\ \mathbf{x}_4 &= \{(0, 2), (1, 2), (2, 1), (3, 2), (4, 3), (5, 3), (6, 3), (7, 3), (8, 3), (9, 3), (10, 3)\} \end{aligned}$$

Para  $i = 1, j = 2$  se tiene  $n_{12}^{(1)} = 1$ , ya que en la primera trayectoria solo se presenta una transición del estado 1 al estado 2, en contraste en la segunda trayectoria  $n_{12}^{(2)} = 3$  ya que se presenta una transición del estado 1 al estado 2 en los instantes del 1 al 2, del 6 al 7 y del 9 al 10. Para la tercera y cuarta trayectoria,  $n_{12}^{(3)} = 2$  y  $n_{12}^{(4)} = 1$ , respectivamente. En este caso, el número de transiciones que se observan del estado 1 al estado 2 está dado por  $n_{12} = 7$ .

Ahora, sea  $\vec{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$ . Ya que las trayectorias son independientes, la función de verosimilitud,  $L(\mathbf{p}_0, \mathbf{P}) = L(\mathbf{p}_0, \mathbf{P}; \vec{\mathbf{x}})$ , está dada por:

$$L(\mathbf{p}_0, \mathbf{P}) = \prod_{a=1}^m P(X_0 = s_0^{(a)}, \dots, X_h = s_h^{(a)}).$$

La propiedad de Markov y la Proposición 1 implican:

$$\begin{aligned} L(\mathbf{p}_0, \mathbf{P}) &= \prod_{a=1}^m P(X_0 = s_0^{(a)}, \dots, X_h = s_h^{(a)}) \\ &= \prod_{a=1}^m p_{s_0^{(a)}} \prod_{i=1}^s \prod_{j=1}^s p_{ij}^{n_{ij}^{(a)}} \\ &= \prod_{k=1}^s p_k^{n_k^0} \prod_{i=1}^s \prod_{j=1}^s p_{ij}^{n_{ij}}, \end{aligned} \quad (3)$$

donde  $n_k^0$  es el número de trayectorias en la muestra que inician en el estado  $k$ . Nótese que, por el Criterio de Factorización [7, Teorema 6.2.6], los estadísticos  $n_{ij}$  son suficientes para los parámetros de la cadena de Markov (la matriz de transición). Ahora bien, de la ecuación (3) se sigue que la función logverosimilitud está dada por:

$$\begin{aligned} l(\mathbf{p}_0, \mathbf{P}) &= \log L(\mathbf{p}_0, \mathbf{P}) \\ &= \sum_{k=1}^s \log p_k^{n_k^0} + \sum_{i=1}^s \sum_{j=1}^s \log p_{ij}^{n_{ij}} \\ &= \sum_{k=1}^s \log p_k^{n_k^0} + \sum_{i=1}^s \log L_i(\mathbf{P}). \end{aligned} \quad (4)$$

donde

$$L_i(\mathbf{P}) = \prod_{j=1}^s p_{ij}^{n_{ij}}. \quad (5)$$

Como puede notarse,  $L_i(\mathbf{P})$  depende solo de los elementos de la  $i$ -ésima fila de la matriz de transición  $\mathbf{P}$ . Esta observación permite establecer lo mencionado en [9]: las probabilidades de transición de la  $i$ -ésima fila y el número de transiciones desde este estado, pueden visualizarse como los parámetros de una distribución multinomial y con base en esto el proceso de maximización de la función de verosimilitud se restringe a maximizar la función de verosimilitud para muestras de distribuciones multinomiales independientes. De manera precisa si tenemos las observaciones  $(n_{i1}, n_{i2}, \dots, n_{is})$  y consideramos una distribución multinomial con parámetros  $n_i, p_{i1}, \dots, p_{is}$ , donde  $p_{ij}$  son las probabilidades de transición desde el estado  $i$  al estado  $j$  y  $n_i$  es el número de transiciones que se realizan en la muestra desde este estado, entonces para hallar los EMV para  $p_{ij}$  se maximiza la función dada en la ecuación (5). Al realizar lo anterior se establece que los estimadores de máxima verosimilitud para los parámetros  $p_{ij}$  están dados por:

$$\hat{p}_{ij} = \frac{n_{ij}}{n_i}. \quad (6)$$

En [14], vía multiplicadores de Lagrange se llega a la misma ecuación (6). El procedimiento planteado en [14] establece que si se desea maximizar  $l(\mathbf{p}_0, \mathbf{P})$  bajo las restricciones:

$$\sum_{i=1}^s p_i = 1, \quad \sum_{j=1}^s p_{ij} = 1, \quad (7)$$

por lo previamente establecido, esto se consigue utilizando multiplicadores de Lagrange, derivando término a término, respecto a  $p_{ij}$ , las funciones

$$\log L_i(\mathbf{P}) + \lambda \left( \sum_{j=1}^s p_{ij} - 1 \right) = \sum_{j=1}^s n_{ij} \log p_{ij} + \lambda \left( \sum_{j=1}^s p_{ij} - 1 \right),$$

e igualando a cero las derivadas. Realizando los cálculos se establece que  $p_{ij} = -n_{ij}/\lambda$ . Utilizando la segunda restricción de las ecuaciones dadas en (7) y la definición de  $n_i$ , se obtiene

la ecuación (6). El mismo procedimiento, asegura los EMV de los parámetros desconocidos de la distribución inicial, a saber,

$$\hat{p}_k = \frac{n_k^0}{m},$$

donde  $n_k^0$  es el número de trayectorias en la muestra que inician en el estado  $k$ . Estos estimadores también fueron obtenidos en [1] y es posible obtenerlos siguiendo la misma idea mencionada en [9], visualizando los parámetros  $m, p_1, \dots, p_s$  como los correspondientes en una distribución multinomial.

Durante el proceso de la obtención de los EMV, se requiere de la diferenciación de la función logverosimilitud obtenida en la ecuación (4). En los procedimientos mencionados previamente, se asume que la distribución inicial no tiene información acerca de los parámetros  $p_{ij}$ , por lo que el primer término del lado derecho de la ecuación (4) no afecta a los EMV dados en la ecuación (6). También, ese primer término del lado derecho de la ecuación (4) puede ser ignorado según [3, p. 53], en el caso de que el tamaño de muestra  $m$  sea muy grande.

El siguiente ejemplo presenta un modelo de cadena de Markov para el deterioro cognitivo de pacientes hasta llegar a la enfermedad de Alzheimer. En este ejemplo se ilustra como sería la muestra aleatoria para el modelo según lo establecido en la ecuación (2).

**Ejemplo 4.** *En [15] se presenta un estudio realizado en una comunidad de China. En este estudio el estado cognitivo de 6,192 sujetos mayores a 65 años es monitoreado cada 6 meses durante 5 años. La variable aleatoria  $X_n$  es el estado de salud del paciente en el semestre  $n$ , siendo  $X_0$  el estado de salud del paciente al comienzo del estudio. El espacio de estados es  $S = \{1, 2, 3\}$ , donde:*

- 1 - Deterioro cognitivo leve
- 2 - Deterioro global
- 3 - Diagnóstico Alzheimer

*Es importante mencionar que si el paciente es diagnosticado con Alzheimer, al ser una enfermedad progresiva, el paciente se mantendrá con ese diagnóstico a lo largo del tiempo, es decir, el estado 3 se considera un estado absorbente. La propiedad de Markov se satisface ya que el resultado de la evaluación depende únicamente del estado en la evaluación anterior más reciente y no depende del estado en otros exámenes anteriores.*

*El monitoreo del estado de salud de cada paciente durante los 5 años se considera como una trayectoria de longitud  $h = 10$ . Por ejemplo, el registro de la evolución cognitiva de un paciente puede ser:*

$$\{(0, 1), (1, 1), (2, 2), (3, 2), (4, 2), (5, 1), (6, 2), (7, 3), (8, 3), (9, 3), (10, 3)\}.$$

*Gráficamente este registro se muestra en la Figura 2.*

*Se puede observar que el paciente ingresó al estudio con un deterioro cognitivo leve y en el primer semestre se mantuvo en un mismo estado de salud, sin embargo, en el semestre 2*

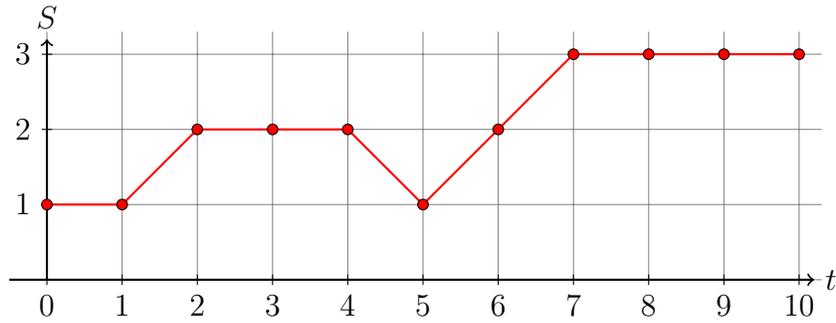


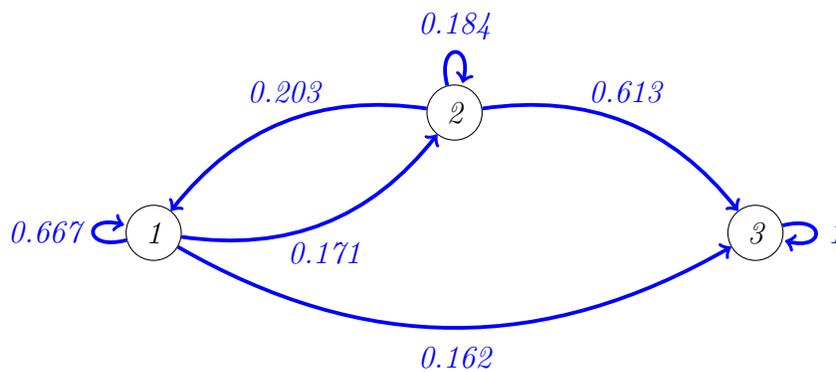
Figura 2: Trayectoria del paciente

*cambió a un estado de deterioro global y se mantuvo en el mismo estado de salud hasta el cuarto semestre, al quinto semestre cambió a un estado de deterioro cognitivo leve, en el sexto semestre se encontró en un deterioro global y en el séptimo semestre el paciente fue diagnosticado con Alzheimer y se mantuvo en ese estado durante los semestres restantes.*

*Con base en las 6,192 trayectorias de longitud 10, se calculan los conteos  $n_{ij}^{(a)}$ ,  $n_{ij}$  y  $n_i$ . Utilizando la ecuación (6) se obtiene la matriz de transición estimada  $\hat{\mathbf{P}} = (\hat{p}_{ij})$ :*

$$\hat{\mathbf{P}} = \begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{pmatrix} 0.667 & 0.171 & 0.162 \\ 0.203 & 0.184 & 0.613 \\ 0 & 0 & 1 \end{pmatrix} \end{matrix}.$$

*El diagrama de transición para el modelo está dado por:*



*Del diagrama anterior se puede concluir que si el paciente se encuentra en un estado cognitivo con deterioro leve es más probable que en seis meses permanezca con estado cognitivo leve que ser diagnosticado con Alzheimer. De igual forma, si el paciente se encuentra en un estado cognitivo con deterioro global, es más probable ser diagnosticado con Alzheimer en seis meses, que permanecer en este estado o tener una mejoría a un estado cognitivo con deterioro leve.*

### 2.1.2 Estimación por intervalos

La distribución de cada estimador  $\hat{p}_{ij}$  no se conoce, sin embargo, en [1] se demuestra que el vector  $(\sqrt{n_i}(\hat{p}_{ij} - p_{ij}))_{j=1,\dots,s}$ , para cada  $i$  fijo, es asintóticamente normal (cuando  $n_i \rightarrow \infty$ ) con vector de medias 0 y matriz de varianzas y covarianzas dada por:

$$\Sigma = \begin{pmatrix} p_{i1}(1 - p_{i1}) & -p_{i1}p_{i2} & \cdots & -p_{i1}p_{is} \\ -p_{i2}p_{i1} & p_{i2}(1 - p_{i2}) & \cdots & -p_{i2}p_{is} \\ \vdots & \vdots & \cdots & \vdots \\ -p_{is}p_{i1} & -p_{is}p_{i2} & \cdots & p_{is}(1 - p_{is}) \end{pmatrix}.$$

Lo anterior permite el cálculo de intervalos asintóticos de confianza para los parámetros  $p_{ij}$ .

Las probabilidades de transición de la  $i$ -ésima fila de la matriz de transición  $\mathbf{P}$  y el número de transiciones desde el estado  $i$ , pueden visualizarse como los parámetros de una distribución multinomial. Esto permite que los intervalos de confianza para las probabilidades de transición en una cadena de Markov se calculen considerando la distribución multinomial. Por lo tanto, para un valor fijo  $i \in S$  se calculan  $s$  intervalos de confianza simultáneos para  $p_{i1}, p_{i2}, \dots$ , y  $p_{is}$ . Los intervalos de confianza simultáneos para las probabilidades de transición  $(p_{i1}, p_{i2}, \dots, p_{is})$ , para  $i \in S$  y con un nivel de confianza del  $(1 - \alpha)100\%$  se denotan por  $I_\alpha = (I_\alpha^1, I_\alpha^2, \dots, I_\alpha^s)$ .

El construir intervalos de confianza simultáneos para más de dos parámetros ha sido un tema estudiado desde hace mucho tiempo, y casi siempre se tiene más de una solución, dependiendo del enfoque utilizado, siempre cuidando que se tenga al menos el nivel de confianza establecido al principio del análisis. En [21] se presenta una revisión de la literatura y se ilustran distintas maneras de calcular intervalos de confianza simultáneos para proporciones multinomiales asintóticos ( $n_i \rightarrow \infty$ ). A continuación se presentan cuatro maneras de calcular estos intervalos de confianza. Es conveniente aclarar que estos intervalos asintóticamente tienen el nivel de confianza del  $100(1 - \alpha)\%$  y en [21] se calcula el nivel de confianza exacto para muestras finitas.

En [11] citado en [21], el intervalo de confianza  $I_\alpha^j$  para  $j = 1, \dots, s$  está dado por:

$$I_\alpha^j = \left( \hat{p}_{ij} - (\chi_{s-1, \alpha}^2)^{1/2} \left[ \frac{\hat{p}_{ij}(1 - \hat{p}_{ij})}{n_i} \right]^{1/2}, \hat{p}_{ij} + (\chi_{s-1, \alpha}^2)^{1/2} \left[ \frac{\hat{p}_{ij}(1 - \hat{p}_{ij})}{n_i} \right]^{1/2} \right). \quad (8)$$

En [12] citado en [21] se consideran los intervalos de Bonferroni  $I_\alpha^j$  y están dados por:

$$I_\alpha^j = \left( \hat{p}_{ij} - z_{\frac{\alpha}{2s}} \left[ \frac{\hat{p}_{ij}(1 - \hat{p}_{ij})}{n_i} \right]^{1/2}, \hat{p}_{ij} + z_{\frac{\alpha}{2s}} \left[ \frac{\hat{p}_{ij}(1 - \hat{p}_{ij})}{n_i} \right]^{1/2} \right). \quad (9)$$

Los intervalos de confianza simultáneos  $I_\alpha^j$  considerados por [19] citado en [21] están dados por:

$$I_\alpha^j = \left( \frac{c + 2n_i \cdot \hat{p}_{ij} - \sqrt{c^2 + 4c \cdot n_i(1 - \hat{p}_{ij})\hat{p}_{ij}}}{2(c + n_i)}, \frac{c + 2n_i \cdot \hat{p}_{ij} + \sqrt{c^2 + 4c \cdot n_i(1 - \hat{p}_{ij})\hat{p}_{ij}}}{2(c + n_i)} \right), \quad (10)$$

donde  $c = \chi_{s-1, \alpha}^2$ .

En [10] citado en [21], el intervalo de confianza  $I_\alpha^j$  para  $j = 1, \dots, s$  está dado por:

$$I_\alpha^j = \left( \hat{p}_{ij} - \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n_i}}, \hat{p}_{ij} + \frac{z_{\frac{\alpha}{2}}}{2\sqrt{n_i}} \right). \quad (11)$$

El siguiente ejemplo consiste en un estudio en el que se modela la progresión de la diabetes en un conjunto de 6,814 pacientes en Estados Unidos. En este ejemplo, complementamos la estimación puntual de las probabilidades de transición con los intervalos de confianza calculados a través de las ecuaciones (8), (9), (10) y (11).

**Ejemplo 5.** En [17] se estudia la diabetes en un conjunto de pacientes mediante una cadena de Markov, el espacio de estados en el modelo de Markov está basado en la clasificación del test de azúcar en la sangre (FBS), es decir,  $S = \{1, 2, 3, 4\}$ , donde:

- 1 - Normal (FBS < 100 mg/dl)
- 2 - Alteración de la glucosa I (100 mg/dl < FBS < 110 mg/dl)
- 3 - Alteración de la glucosa II (110 mg/dl < FBS < 126 mg/dl)
- 4 - Diabetes

$$\mathbf{N} = \begin{array}{c} \begin{array}{cccc} & 1 & 2 & 3 & 4 \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} n_{11} & n_{12} & n_{13} & n_{14} \\ n_{21} & n_{22} & n_{23} & n_{24} \\ n_{31} & n_{32} & n_{33} & n_{34} \\ n_{41} & n_{42} & n_{43} & n_{44} \end{pmatrix} & \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 12,149 & 1708 & 320 & 630 \\ 1150 & 738 & 346 & 249 \\ 202 & 230 & 249 & 261 \\ 452 & 126 & 100 & 2339 \end{pmatrix} \end{array} \end{array}, \quad (12)$$

$$\mathbf{P} = \begin{array}{c} \begin{array}{cccc} & 1 & 2 & 3 & 4 \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0.82 & 0.11 & 0.02 & 0.04 \\ 0.46 & 0.29 & 0.13 & 0.10 \\ 0.21 & 0.24 & 0.26 & 0.27 \\ 0.14 & 0.04 & 0.03 & 0.77 \end{pmatrix} \end{array} \end{array}. \quad (13)$$

Con base en las matrices (12) y (13) se calculan para las probabilidades de transición  $p_{ij}$  sus respectivos intervalos de confianza simultáneos con un nivel de confianza al 95%. Los resultados se muestran a continuación.

|         | $p_{i1}$       | $p_{i2}$       | $p_{i3}$       | $p_{i4}$       |
|---------|----------------|----------------|----------------|----------------|
| $i = 1$ | [0.811, 0.828] | [0.102, 0.117] | [0.016, 0.023] | [0.035, 0.044] |
| $i = 2$ | [0.432, 0.487] | [0.264, 0.315] | [0.111, 0.148] | [0.083, 0.116] |
| $i = 3$ | [0.172, 0.247] | [0.201, 0.278] | [0.220, 0.299] | [0.229, 0.310] |
| $i = 4$ | [0.122, 0.157] | [0.030, 0.049] | [0.021, 0.038] | [0.748, 0.791] |

Tabla 1: Intervalos de confianza simultáneos basados en [11] para  $p_{ij}$ .

|         | $p_{i1}$       | $p_{i2}$       | $p_{i3}$       | $p_{i4}$       |
|---------|----------------|----------------|----------------|----------------|
| $i = 1$ | [0.812, 0.827] | [0.103, 0.116] | [0.017, 0.022] | [0.035, 0.044] |
| $i = 2$ | [0.435, 0.484] | [0.267, 0.312] | [0.113, 0.146] | [0.084, 0.115] |
| $i = 3$ | [0.176, 0.243] | [0.205, 0.274] | [0.224, 0.295] | [0.233, 0.306] |
| $i = 4$ | [0.124, 0.155] | [0.031, 0.048] | [0.022, 0.037] | [0.750, 0.789] |

Tabla 2: Intervalos de confianza simultáneos basados en [12] para  $p_{ij}$ .

|         | $p_{i1}$       | $p_{i2}$       | $p_{i3}$       | $p_{i4}$       |
|---------|----------------|----------------|----------------|----------------|
| $i = 1$ | [0.811, 0.828] | [0.103, 0.117] | [0.017, 0.023] | [0.035, 0.044] |
| $i = 2$ | [0.432, 0.488] | [0.265, 0.316] | [0.112, 0.150] | [0.084, 0.118] |
| $i = 3$ | [0.175, 0.249] | [0.203, 0.280] | [0.222, 0.301] | [0.231, 0.312] |
| $i = 4$ | [0.123, 0.158] | [0.031, 0.051] | [0.022, 0.039] | [0.747, 0.790] |

Tabla 3: Intervalos de confianza simultáneos basados en [19] para  $p_{ij}$ .

|         | $p_{i1}$       | $p_{i2}$       | $p_{i3}$       | $p_{i4}$       |
|---------|----------------|----------------|----------------|----------------|
| $i = 1$ | [0.811, 0.828] | [0.101, 0.118] | [0.011, 0.028] | [0.031, 0.048] |
| $i = 2$ | [0.440, 0.479] | [0.270, 0.309] | [0.110, 0.149] | [0.080, 0.119] |
| $i = 3$ | [0.178, 0.241] | [0.208, 0.271] | [0.228, 0.291] | [0.238, 0.301] |
| $i = 4$ | [0.122, 0.157] | [0.022, 0.057] | [0.012, 0.047] | [0.752, 0.787] |

Tabla 4: Intervalos de confianza simultáneos basados en [10] para  $p_{ij}$ .

Un valor de interés es conocer la probabilidad de que un paciente con un nivel de azúcar normal se mantenga en ese estado al transcurrir una unidad de tiempo, es decir, interesa conocer  $p_{11}$ . De acuerdo con la Tabla 1, con un nivel de confianza al 95 %, la probabilidad  $p_{11}$  se encuentra entre 0.811 y 0.828. Este intervalo es igual al presentado en las Tablas 3 y 4. Note que el intervalo obtenido mediante la aproximación dada en [12] es de menor longitud que los otros (ver Tabla 2) y que los 4 intervalos presentan resultados similares.

## 2.2 Prueba de homogeneidad

En esta sección se presenta una prueba de hipótesis basada en la razón de verosimilitud y contrasta la hipótesis: una cadena de Markov es homogénea; contra la hipótesis: la cadena de Markov no es homogénea. Para el planteamiento de la prueba, es de suma importancia presentar algunos elementos de las cadenas de Markov no homogéneas: función de verosimilitud, estimadores de las probabilidades de transición y el número de probabilidades de transición a estimar.

Los supuestos para la prueba de hipótesis son: todas las trayectorias que conforman la muestra deben tener el mismo estado inicial y cada trayectoria tiene una longitud  $h$ .

Estos supuestos permiten plantear la función de verosimilitud y calcular los estimadores de máxima verosimilitud que son primordiales para la prueba de hipótesis. Las probabilidades de transición en una cadena de Markov no homogénea dependen del tiempo y se denotan como  $p_{ij}(t) = P(X_t = j \mid X_{t-1} = i)$ , donde la probabilidad  $p_{ij}(t)$  es la probabilidad de transitar del estado  $i$  en el tiempo  $t - 1$  al estado  $j$  en el tiempo  $t$ . Es importante considerar el tiempo de observación en la muestra ya que las probabilidades de transición  $p_{ij}(t)$  dependen de él. Por cada tiempo  $t \in \{1, \dots, h\}$  se pueden organizar las probabilidades de transición  $p_{ij}(t)$  en una matriz la cual se denota por  $\mathbf{P}(t) = (p_{ij}(t))$  y en cada matriz  $\mathbf{P}(t)$  se estiman  $s(s - 1)$  probabilidades de transición. Por lo tanto, se estiman en total  $s(s - 1)h$  probabilidades de transición.

El cálculo de la función de verosimilitud para una cadena de Markov no homogénea es análogo al de una cadena de Markov homogénea, esto es, considerando  $\vec{\mathbf{x}} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$ ,  $\mathbf{t} = (0, 1, \dots, h)$  y la independencia de las trayectorias, se sigue que la función de verosimilitud,  $L(\mathbf{p}_0, \mathbf{P}(\mathbf{t})) = L(\mathbf{p}_0, \mathbf{P}(\mathbf{t}); \vec{\mathbf{x}})$ , está dada por:

$$\begin{aligned} L(\mathbf{p}_0, \mathbf{P}(\mathbf{t})) &= \prod_{a=1}^m P(X_0 = s_0^{(a)}, \dots, X_h = s_h^{(a)}) \\ &= \prod_{k=1}^s p_k^{n_k^0} \prod_{t=1}^h \prod_{i=1}^s \prod_{j=1}^s p_{ij}(t)^{n_{ij}(t)}, \end{aligned} \quad (14)$$

donde  $n_{ij}(t) = \sum_{a=1}^m n_{ij}^{(a)}(t)$  es el número de veces que se observó en la muestra la transición del estado  $i$  en el tiempo  $t - 1$  al estado  $j$  en el tiempo  $t$ .

Los estimadores de máxima verosimilitud para una cadena de Markov no homogénea fueron calculados en [1] y están dados por:

$$\hat{p}_{ij}(t) = \frac{n_{ij}(t)}{n_i(t-1)}, \quad (15)$$

donde  $n_i(t-1) = \sum_{j=1}^s n_{ij}(t)$ ,  $t = 1, 2, \dots, h$ , es decir,  $n_i(t-1)$  es el número de transiciones que parten del estado  $i$  en el tiempo  $t - 1$ .

La siguiente prueba de hipótesis está basada en la razón de verosimilitud. Esta prueba se puede encontrar en [1].

### Hipótesis:

$$H_0 : p_{ij}(t) = p_{ij}, \quad \text{para toda } t \in \{1, 2, \dots, h\}, \quad \text{para } i, j \in S.$$

vs

$$H_1 : p_{ij}(t_0) \neq p_{ij}(t_1), \quad \text{para algunos } t_0, t_1 \in \{1, \dots, h\}, i, j \in S.$$

En la hipótesis nula se tiene que la cadena de Markov es homogénea, es decir, las probabilidades de transición en un paso no dependen del tiempo  $t$ , por el contrario en la hipótesis

alternativa sí se establece esta dependencia, lo que implica que la cadena de Markov no es homogénea.

El estadístico de prueba se obtiene a partir de considerar la función de verosimilitud presentada en la ecuación (14) y los estimadores  $\hat{p}_{ij}$ ,  $\hat{p}_{ij}(t)$ , dados en las ecuaciones (6) y (15), respectivamente. Bajo la hipótesis nula, el máximo de la función de verosimilitud es:

$$\prod_{k=1}^s p_k^{n_k^0} \prod_{t=1}^h \prod_{i=1}^s \prod_{j=1}^s [\hat{p}_{ij}]^{n_{ij}(t)},$$

mientras que bajo la hipótesis alternativa, el máximo de la función de verosimilitud está dado por:

$$\prod_{k=1}^s p_k^{n_k^0} \prod_{t=1}^h \prod_{i=1}^s \prod_{j=1}^s [\hat{p}_{ij}(t)]^{n_{ij}(t)}.$$

Se considera el cociente de las dos últimas cantidades para obtener  $\lambda$ :

$$\lambda = \prod_{t=1}^h \prod_{i=1}^s \prod_{j=1}^s \left( \frac{n_{ij}/n_i}{n_{ij}(t)/n_i(t-1)} \right)^{n_{ij}(t)}. \quad (16)$$

De aquí, el estadístico de prueba es  $U = -2 \log(\lambda)$ , el cual tiene una distribución asintótica ji-cuadrada con  $\rho = s(s-1)(h-1)$  grados de libertad. Los grados de libertad se calculan considerando el número de probabilidades de transición que se estiman en una cadena de Markov no homogénea menos el número de parámetros de una cadena de Markov homogénea. De esta manera, se tiene que el estadístico de prueba y la región de rechazo para la prueba de homogeneidad de una cadena de Markov están dados por:

**Estadístico de prueba:**

$$U = -2 \log(\lambda),$$

donde  $\lambda$  está definida en la ecuación (16).

**Región de rechazo:**

Se rechaza  $H_0$  si  $U > \chi_{\rho, \alpha}^2$ , con  $\rho = s(s-1)(h-1)$ .

En [5] consideran una base de datos dada en [14] para presentar un ejemplo sobre cadenas de Markov. Los autores del presente artículo consideraron esta base de datos para probar la hipótesis de homogeneidad de la cadena de Markov utilizada para modelar la evolución en la problemática descrita en el ejemplo. Esto se proporciona con mayor detalle a continuación.

**Ejemplo 6.** En [14] se presenta un estudio en el que se monitorea el clima durante los días de enero durante 36 años en Snoqualmie Falls, Estados Unidos. El estudio plantea una cadena de Markov  $\{X_n, n \geq 0\}$  con dos estados, donde la variable aleatoria  $X_n$  representa el clima en el día  $n$ , el estado 1 representa un día seco y el estado 2 representa un día mojado.

Los datos conforman 36 trayectorias de longitud  $h = 30$  días. Considerando la muestra aleatoria como en la ecuación (2) y con base en el procedimiento dado en esta sección, se

realiza la prueba de hipótesis de homogeneidad, donde en la hipótesis nula se tiene que la cadena de Markov es homogénea y en la hipótesis alternativa se tiene que la cadena de Markov no es homogénea. El valor del estadístico de prueba obtenido  $U = 32.90$  y con un valor de  $\alpha = 0.05$ , se obtuvo que  $\chi_{58,0.05}^2 = 76.77$ . Como el estadístico de prueba  $U < \chi_{58,0.05}^2$  no existe evidencia suficiente para rechazar la hipótesis nula, por lo que la cadena de Markov que modela el clima de los días de enero en Snoqualmie Falls, Estados Unidos, es homogénea.

### 3. Conclusiones

En este artículo se presentaron algunas aplicaciones del modelo de cadenas de Markov que atienden problemas de diferentes áreas del conocimiento y que muestran los diferentes elementos y propiedades que pueden tener las cadenas de Markov como la distribución inicial, ser homogénea, irreducible y los estados absorbentes.

Cuando el contexto de un problema permite deducir que se cumple la propiedad de Markov, entonces es natural pensar que las Cadenas de Markov serían un modelo adecuado para hacer inferencias y tomar decisiones.

La teoría principal del modelo de cadenas de Markov se centra bajo el supuesto de que la cadena es homogénea, este supuesto en la mayoría de los casos se asume o se justifica con base en el contexto, sin embargo existen pruebas estadísticas que permiten verificar estadísticamente la homogeneidad en las cadenas de Markov como tal es el caso de la prueba de hipótesis presentada en la Sección 2.2 y que está basada en la razón de verosimilitud. Visualizar la muestra como se propone en la Sección 2.1 favorece el conteo de las transiciones en cada tiempo y esto ayuda al cálculo del estadístico de prueba para verificar la homogeneidad mediante la prueba de hipótesis.

La estimación de los parámetros mediante intervalos de confianza son una herramienta útil para complementar la estimación puntual de los estimadores de máxima verosimilitud, por tal razón en este artículo se presentaron varias opciones para estimar los intervalos de confianza en forma simultánea y asintótica para las probabilidades de transición así como el estudio realizado por [17] incorporando los intervalos de confianza simultáneos.

Existe teoría adicional de la mostrada en este documento, sin embargo, el objetivo de este artículo no es saturar de información al lector sino plasmar la teoría principal que se ha desarrollado sobre la inferencia estadística para cadenas de Markov.

### Agradecimientos

Los autores desean expresar su agradecimiento a los revisores anónimos por sus correcciones y comentarios que permitieron obtener una versión mejorada de este manuscrito.

## Referencias

- [1] T. Anderson and L. Goodman, "Statistical inference about Markov chains," *Annals of Mathematical Statistics*, vol. 28, pp. 89–110, 1957.
- [2] P. Avery and D. Henderson, "Fitting Markov chain models to discrete state series such as DNA sequences," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 48, pp. 53–61, 1999.
- [3] I. Basawa and B. P. Rao, *Statistical Inference for Stochastic Process*. London: Academic Press Inc., 1980.
- [4] J. Batún, R. Cool, and H. Pantí, "Analysis of the academic progress of higher education students by means of an Absorbing Markov Chain," *Autonomous University of Yucatan*, vol. (preprint), 2022.
- [5] J. Besag and D. Mondal, "Exact goodness-of-fit tests for Markov chains," *Biometrics*, vol. 69, pp. 488–496, 2013. DOI: 10.1111/biom.12009
- [6] A. Bolivar, C. Notario, and A. Perez, "Modelos de Markov para la trayectoria académica de estudiantes de la UJAT," *Miscelánea Matemática*, vol. 62, pp. 29–43, 2016.
- [7] G. Casella, *Statistical Inference*. United States of America: Duxbury Thompson Learning, 2002.
- [8] B. Craig and P. Sendi, "Estimation of the transition matrix of a discrete-time Markov chain," *Health Economics*, vol. 11, pp. 33–42, 2002. DOI: 10.1002/hec.654
- [9] A. Davidson, *Statistical Models*. New York, United States of America: Cambridge University Press, 2003.
- [10] S. Fitzpatrick and A. Scott, "Quick simultaneous confidence intervals for multinomial proportions," *Journal of the American Statistical Association*, vol. 82, pp. 875–878, 1987.
- [11] R. Gold, "Tests auxiliary to chi square tests in a Markov chain," *The Annals of Mathematical Statistics*, vol. 34, pp. 56–74, 1963.
- [12] L. Goodman, "On simultaneous confidence intervals for multinomial proportions," *Technometrics*, vol. 7, pp. 247–254, 1965.
- [13] P. Gupta, C. Fillmore, G. Jiang, S. Shapira, K. Tao, C. Kupperwasser, and E. Lander, "Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells," *Cells*, vol. 146, no. 4, pp. 633–544, 2011. DOI: 10.1016/j.cell.2011.07.026
- [14] P. Guttorp, *Stochastic Modeling of Scientific Data*, 5th ed. Seattle, United States of America: Springer Science + Business Media Dordrecht, 1995.

- [15] Y. Hong, G. Jian, Z. Li, L. Rui, and Q. Cheng, “Multi-state markov model in outcome of mild cognitive impairments among community elderly residents in mainland china,” *International Psychogeriatrics*, vol. 18, no. 5, pp. 797–804, 2013.
- [16] E. Mohammedelameen, K. Abdellateef, and H. Mubarak, “Using Markov chain to predict by the natural progression of diabetic nephropathy at diabetes and endocrinology hospital,” *IJCSNS International Journal of Computer Science and Network Security*, vol. 19, pp. 132–137, 2019. DOI: 10.3980/j.issn.2222-3959.2015.01.25
- [17] M. Nazaria, S. Nazarib, F. Zayeri, M. Gholampour, and A. Baghbane, “Estimating transition probability of different states of type 2 diabetes and its associated factors using Markov model,” *Primary Care Diabetes*, vol. 12, no. 3, pp. 245–253, 2018. DOI: doi.org/10.1016/j.pcd.2018.01.004
- [18] M. Pinsky and S. Karlin, *An introduction to Stochastic Modeling*. United States of America: Elsevier, 2010.
- [19] C. Quesenberry and D. Hurst, “Large-sample simultaneous confidence intervals for multinomial proportions,” *Primary Care Diabetes*, vol. 6, pp. 191–195, 1964.
- [20] W. Suryaningrat, D. Munandar, A. Maryati, A. S. Abdullah, and B. N. Ruchjana, “Posted prediction in social media base on markov chain model: twitter dataset with covid-19 trends,” *Journal of Physics: Conference Series*, vol. 1722, no. 1, p. 012001, jan 2021. DOI: 10.1088/1742-6596/1722/1/012001
- [21] H. Wang, “Exact confidence coefficients of simultaneous confidence intervals for multinomial proportions,” *Journal of Multivariate Analysis*, vol. 99, no. 5, pp. 896–911, 2008. DOI: <https://doi.org/10.1016/j.jmva.2007.05.003>

**Como citar este artículo:** H. Pantí-Trejo, J. Batún Cutz, R. Cool Padilla, D. Kantún Chim and E. Guerrero Lara, “Aspectos básicos en la Inferencia Estadística para Cadenas de Markov en tiempo discreto”, *Sahuarus. Revista Electrónica de Matemáticas*, vol. 6, no. 1, pp. 30-49, 2022. <https://doi.org/10.36788/sah.v6i1.131>