



"El saber de mis hijos
hará mi grandeza"

SahuarUS

Revista electrónica de Matemáticas

Vol. 6
Número 1

Septiembre 2022 · e-ISSN: 2448-5365



"El saber de mis hijos
hará mi grandeza"

Dra. María Rita Plancarte Martínez

Rectora

Dr. Ramón Enrique Robles Zepeda

Secretario General Académico

Dr. Luis Enrique Riojas Duarte

Secretario General Administrativo

Vicerrectoría de la Unidad Regional Centro

Dra. Diana María Meza Figueroa

Directora de Apoyo a Vinculación y
Difusión

Dr. Rodrigo Meléndrez Amavizca

Director de la División de Ciencias Exactas y
Naturales

Dr. Juan Pablo Soto Barrera

Jefe del Departamento de Matemáticas



Editor Responsable

Dr. Misael Avendaño Camacho

Comité Editorial

Dr. Manuel Adrian Acuña Zegarra

Dra. Carolina Espinoza Villalva

Dra. Carmen Geraldí Higuera Chan

Ing. Aaron Lara Ordoñez

Dra. Gloria Angélica Moreno Durazo

Dr. José Crispín Ruíz Pantaleón

Editores Asociados

Dr. José Luis Cisneros Molina

Instituto de Matemáticas, Unidad Cuernavaca, UNAM

Dr. Xavier Gómez Mont

Centro de Investigaciones en Matemáticas

Dr. Juan Carlos Hernández Gómez

Facultad de Matemáticas, Universidad Autónoma de Guerrero,
Acapulco, Guerrero

Dr. Fernando Antonio Hitt Espinoza

Universidad de Quebec, Montreal, Canada

Dra. Roxana López Cruz

Facultad de Matemáticas, Universidad Nacional Mayor de San Marcos,
Lima, Perú

Dr. Humberto Madrid de la Vega

Centro de Investigación en Matemáticas Aplicadas, Universidad
Autónoma de Coahuila

Dr. Pedro Miramontes Vidal

Facultad de Ciencias, Universidad Nacional Autónoma de México

Dr. Carlos Gabriel Pacheco González

Centro de Investigación y de Estudios Avanzados

Dra. Sandra Evely Parada Rico

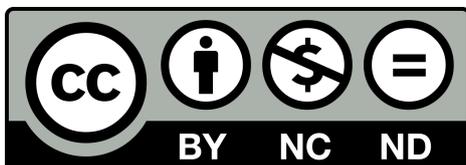
Universidad Industrial de Santander, Bucaramanga, Colombia

Dr. José Antonio Vallejo Rodríguez

Universidad Autónoma de San Luis Potosí

SAHUARUS. REVISTA ELECTRÓNICA DE MATEMÁTICAS, volumen 6, número 1, septiembre 2022 - marzo 2023, es una publicación semestral, editada por la Universidad de Sonora, a través del Departamento de Matemáticas. Blvd. Luis Encinas y Rosales S/N, colonia Centro, Hermosillo, Sonora, México. C.P. 83000. Tel. (662) 2592155. Página web: sahuarus.unison.mx. Correo electrónico: sahuarus@unison.mx. Editor responsable: Misael Avendaño Camacho. Reserva de Derechos al Uso Exclusivo No. **04-2017-062817342200-203**, e-ISSN: 2448-5365, ambos otorgados por el Instituto Nacional de Derechos de Autor.

Los artículos publicados por [Sahuarus. Revista Electrónica de Matemáticas](http://Sahuarus.RevistaElectronica.deMatematicas) se distribuye bajo una [Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional](http://LicenciaCreativeCommonsAtribucion-NoComercial-SinDerivadas4.0Internacional), la cual permite la distribución y el uso del material publicado citando la fuente de la que proviene, prohíbe la modificación y el uso con fines comerciales.



Índice General

Un viaje con velocidad aleatoria y procesos estocásticos ergódicos Juan Ruiz de Chávez & Evgueni Gordienko	1 - 15
Border ideals: un algoritmo de pertenencia Petra Rubi Pantaleón Mondragón	16 - 29
Aspectos básicos en la Inferencia Estadística para Cadenas de Markov en tiempo discreto Henry Pantí, José Luis Batún, Rubén Cool, Diódora Kantún Chim & Ernesto Guerrero Lara	30 - 49

Un viaje con velocidad aleatoria y procesos estocásticos ergódicos

E. Gordienko¹ y J. Ruiz de Chávez²,

Departamento de Matemáticas Universidad Autónoma Metropolitana-I
¹gord@xanum.uam.mx, ²jrch@xanum.uam.mx

Resumen

El objetivo del artículo es doble:

- *Dar un ejemplo donde se muestra la importancia de la selección del modelo adecuado incluso en fenómenos estocásticos muy sencillos;*
- *Dar una imagen preliminar de algunos resultados de Teoría Ergódica de procesos estacionarios y markovianos.*

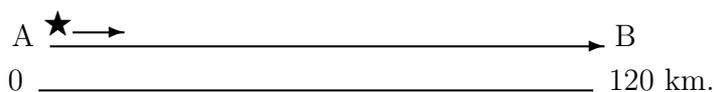
Palabras Clave: Ergodicidad, Proceso estocástico, Velocidad aleatoria.

DOI: 10.36788/sah.v6i1.126

Recibido: 23 de abril de 2022.

Aceptado: 20 de junio de 2022.

1. Un “modelo” de viaje con velocidad aleatoria



Denotemos por V , la variable aleatoria (abreviación: **v.a.**) que indica la velocidad del objeto \star (por ejemplo un coche). Supongamos que la velocidad promedio (valor esperado de V) es

$$\mathbb{E}V = 60 \text{ km/h.}$$

Sea T la v.a. que representa el tiempo utilizado para viajar de A a B . ¿Cuál es el tiempo esperado del viaje $\mathbb{E}T$? Pareciera ser:

$$\mathbb{E}T = \frac{120}{60} = 2 \text{ (horas)}. \quad (1)$$

¿Es correcto el resultado en (1)?

Consideremos, por ejemplo, el caso en donde la v.a. V (la velocidad) tiene una densidad de probabilidad f_V que es continua en el intervalo $[0, \infty)$. (Notemos que $f_V = 0$ para $x < 0$, pues $V \geq 0$.) Además, supondremos que:

$$\lim_{x \rightarrow 0^+} f_V(x) = \alpha > 0, \quad (\text{el límite por la derecha}), \text{ y también} \quad (2)$$

$$\mathbb{E}V = \int_0^{\infty} x f_V(x) dx = 60 \quad (\text{km/h}). \quad (3)$$

Las condiciones (2) y (3) se cumplen, por ejemplo, si f_V es uniforme en el intervalo $[0, 120]$, o cuando la v.a. V es exponencial de parámetro $1/60$.

Vamos, ahora, a tratar de comprobar la igualdad (1). De los cursos de Física elementales que se estudian en la preparatoria sabemos que el tiempo T que se ocupa para recorrer la distancia S , con la velocidad V , es

$$T = \frac{S}{V} = \frac{120}{V}. \quad (4)$$

Por otro lado, si la v.a. V tiene la densidad f_V y $T = g(V) = \frac{120}{V}$, entonces el *tiempo promedio de viaje* es

$$\mathbb{E}T = \int_0^{\infty} g(x) f_V(x) dx = \int_0^{\infty} \frac{120}{x} f_V(x) dx. \quad (5)$$

De la continuidad de f_V y (2) existe un número $\gamma > 0$ tal que

$$f_V(x) \geq \frac{\alpha}{2} \quad \text{para toda } x \in [0, \gamma]. \quad (6)$$

De (5) y (6) se sigue que

$$\mathbb{E}T \geq 120 \int_0^{\gamma} \frac{1}{x} f_V(x) dx \geq 60\alpha \int_0^{\gamma} \frac{1}{x} dx. \quad (7)$$

Como la función $1/x$ no es acotada en cero, la última integral en (7) es impropia, entonces se calcula como sigue. Para $0 < \varepsilon < \gamma$,

$$\int_0^{\gamma} \frac{dx}{x} = \lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^{\gamma} \frac{dx}{x} = \lim_{\varepsilon \rightarrow 0} [\ln(\gamma) - \ln(\varepsilon)] = \infty, \quad (8)$$

ya que $\ln(\varepsilon) \rightarrow -\infty$ cuando $\varepsilon \rightarrow 0^+$. De (7) y (8) se obtiene que

$$\mathbb{E}T = \infty. \quad (9)$$

Por tanto, en lugar de la igualdad “intuitiva” (1), hemos obtenido un **resultado absurdo** (9), que significa que en promedio un viaje va a durar un tiempo infinito. Como no se han

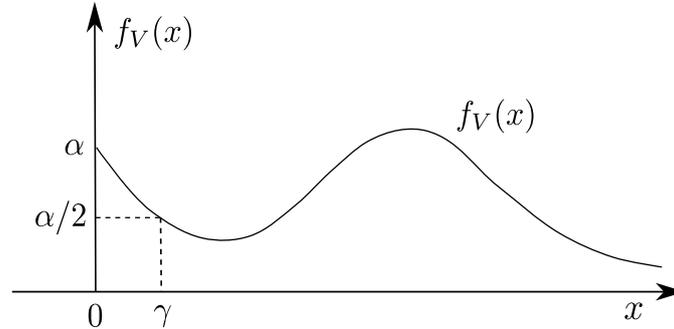
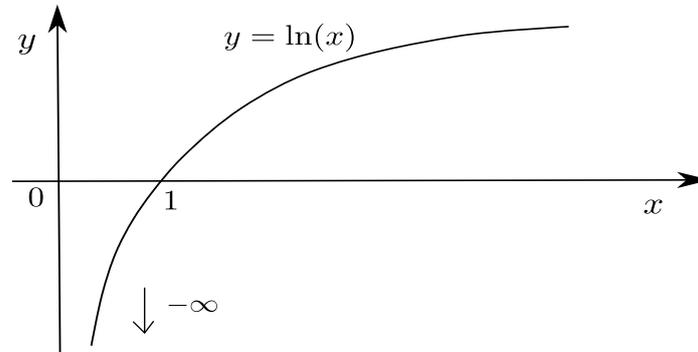
Figura 1: Densidad de V .

Figura 2: Gráfica del Logaritmo.

cometido errores matemáticos, hay que concluir que el modelo anterior de “viaje con velocidad aleatoria” *no es adecuado a la realidad*. La fuente del desajuste es *el uso de la ecuación (4)*, que es cierta solamente cuando durante todo el movimiento la velocidad V es *constante*, es decir, no depende del tiempo t que va transcurriendo.

“El resultado” (9) es una consecuencia de (4) y el hecho de que debido a (2) la velocidad V puede tomar valores arbitrariamente cercanos a cero. Por ejemplo, si un valor de V es 1 cm/h, entonces de (4), el viaje va a durar 12,000,000 horas.

En realidad, la velocidad V es una v.a. que depende del tiempo.

Imaginemos que el movimiento (o viaje) comienza en algún momento t_0 , que vamos a suponer por simplicidad que $t_0 = 0$. Denotemos por $t \in [0, T]$ *el tiempo que transcurre durante el viaje*.

Para movimientos reales (de un coche, por ejemplo) en cada instante t la velocidad es una *variable aleatoria*, es decir V depende de algunos “factores aleatorios” que denotamos por ω . Sin embargo, esta v.a. depende del tiempo t , es decir, $V = V(t, \omega)$. En otras palabras, la velocidad es *un proceso estocástico*.

2. Procesos estocásticos. Promedios con respecto a factores aleatorios y el tiempo

Definición 1 Sea $X(t, \omega)$ una colección de v.a.'s, definidas en un espacio de probabilidad $(\Omega, \mathfrak{F}, P)$, que dependen del tiempo $t \in [0, \infty)$ (ω es el parámetro que es responsable de la "aleatoriedad", $\omega \in \Omega$). La colección $\{X(t, \cdot)\}_{t \geq 0}$ se llama proceso estocástico o aleatorio en tiempo continuo (se denota por $X(t)$). Cuando $t = n = 0, 1, 2, \dots$, se dice que el proceso $X(n, \omega)$ es en tiempo discreto (se denota por X_n).

Ejemplo 1 (a) $X(n, \omega) = X_n$, $n = 1, 2, \dots$, donde las X_n , $n = 1, 2, \dots$, son v.a.'s independientes e idénticamente distribuidas.

(b) El movimiento browniano $B(t) \equiv B_t$. ($t \geq 0$) el cual es uno de los procesos más estudiados. Fue descubierto por Robert Brown en 1823, posteriormente A. Einstein publicó tres artículos sobre el mismo tema en los años 1905-1906 y un libro en 1926 (1). Se tienen registrados hasta la fecha en Mathematical Reviews más de 5000 artículos y 60 libros. Ha sido utilizado para modelar fenómenos físicos como una idealización matemática del movimiento real de partículas en suspensión en un líquido, y otro ejemplo es la descripción de las fluctuaciones de los precios de las acciones en la bolsa de valores, esta última hecha en 1900 por Louis Bachelier, alumno de Henry Poincaré. Para terminar, citaremos la observación de Jean Perrin (6) Sec. 13, trabajo que le hizo acreedor al premio nobel de Física en 1926 "El movimiento browniano es un caso en donde resulta natural pensar en las funciones continuas sin derivadas que los matemáticos han imaginado y que se veían como meras curiosidades matemáticas". Para una revisión más exhaustiva hasta los años 1950 se puede consultar el artículo de J. P. Kahane (4).

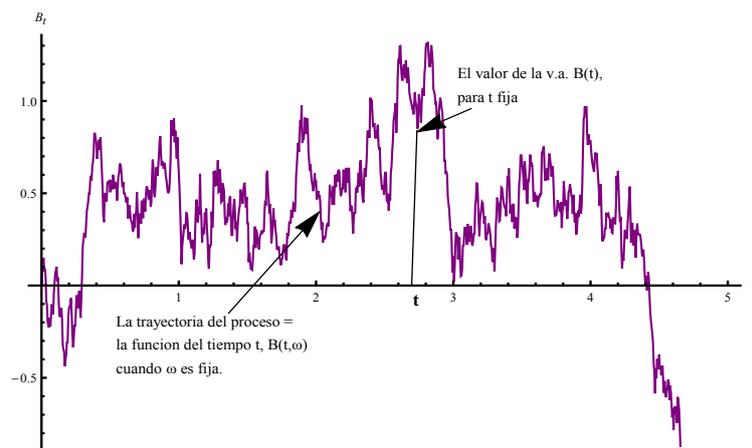


Figura 3: Esta figura muestra una simulación en Mathematica de una trayectoria del movimiento browniano.

Sea $X(t)$ un proceso estocástico. Al fijar un $t \geq 0$, $X(t)$ es una v.a. con *promedio* (valor esperado) $\mathbb{E}X(t)$. Aquí se supone que dicho promedio es finito, y, por ejemplo, cuando $X(t)$ tiene una densidad f_t ,

$$\mathbb{E}X(t) = \int_{-\infty}^{\infty} x f_t(x) dx, \quad t \geq 0. \quad (10)$$

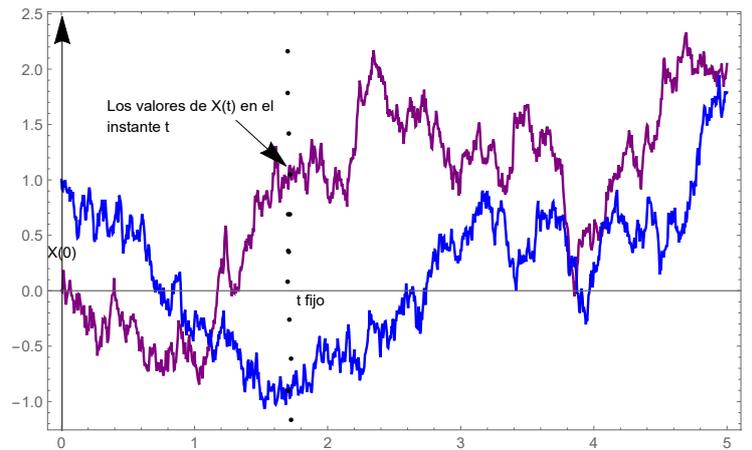


Figura 4: Esta figura muestra los valores del proceso en el instante t (y dos trayectorias).

Al número $\mathbb{E}X(t)$ en (10) vamos a llamarlo *el promedio* de $X(t)$ con respecto a *factores aleatorios* $\omega \in \Omega$, en el instante t . Este nombre se debe a que (10) es un caso particular de la siguiente definición de promedio (consultar por ejemplo (8)):

$$\mathbb{E}X(t) = \int_{\Omega} X(t, \omega) dP(\omega), \quad (t \text{ es fijo}). \quad (11)$$

Al fijar un “factor aleatorio” $\omega \in \Omega$, obtenemos una función del tiempo $X(t, \omega) \equiv X(t)$, que se llama **trayectoria del proceso** $X(t)$.

En la Figura 4 están trazadas dos trayectorias (para dos diferentes ω 's) del proceso $X(t)$.

De nuevo, fijemos un $\omega \in \Omega$ y observemos la trayectoria $X(t) = X(t, \omega)$ (definida para todas las $t \geq 0$).

Sea $T > 0$ un número fijo. *El promedio* de $X(t)$ con respecto al tiempo en el intervalo $[0, T]$ es la siguiente v.a.:

$$\frac{1}{T} \int_0^T X(t) dt.$$

Esta es en efecto una v.a. puesto que la trayectoria $X(t)$ depende de $\omega \in \Omega$.

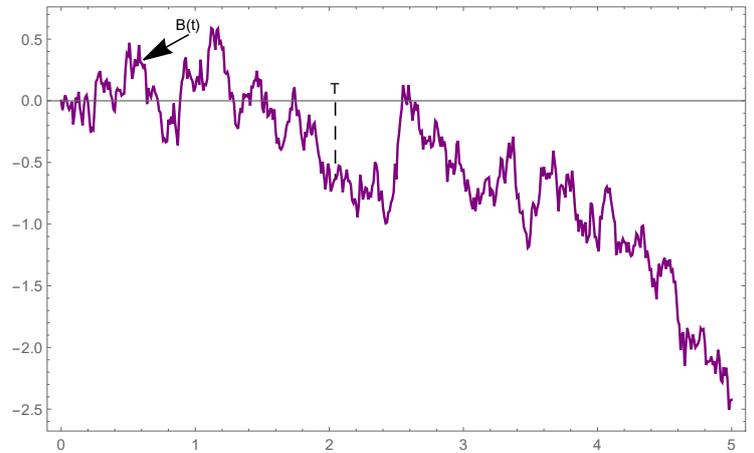


Figura 5: Esta figura muestra una simulación en Mathematica de una trayectoria del movimiento browniano.

Al suponer que el siguiente límite existe

$$\mu_X \stackrel{\text{def}}{=} \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T X(t) dt, \quad (12)$$

la v.a. μ_X es llamada *promedio* de $X(t)$ con respecto al tiempo. De entre algunas definiciones de ergodicidad diferentes vamos, por el momento, a utilizar la siguiente.

Definición 2 Se dice que un proceso estocástico a tiempo continuo $X(t)$, $t \geq 0$ es ergódico, si

- (a) existe una constante m_X tal que $\mathbb{E}X(t) = \mathbb{E}X(0) = m_X$ para toda $t \geq 0$;
- (b) con probabilidad 1, el límite en (12) existe y además,

$$\mu_X = m_X. \quad (13)$$

En otras palabras, el promedio (con respecto a ω) en (11) no depende del tiempo t y dicho promedio coincide con el promedio tomado con respecto al tiempo μ_X . Como consecuencia el último promedio μ_X debe ser una v.a. constante (con probabilidad 1).

En la Sección 5 se presentarán condiciones que son suficientes para que un proceso sea ergódico. Ahora damos un ejemplo muy sencillo.

Ejemplo 2 Sea $X(t) = \text{sen}(\xi t)$, $t \geq 0$ donde ξ es una v.a. con densidad uniforme en el intervalo $[-1, 1]$. Entonces para $t > 0$

$$\begin{aligned} \mathbb{E}X(t) &= \frac{1}{2} \int_{-1}^1 \text{sen}(xt) dx = \frac{1}{2t} \int_{-t}^t \text{sen}(y) dy \\ &= -\frac{1}{2t} [\cos(t) - \cos(-t)] = 0. \end{aligned}$$

Por otro lado, es evidente que para cada $\xi = x$, $\frac{1}{T} \int_0^T \text{sen}(xt) dt \rightarrow 0$ cuando $T \rightarrow \infty$.

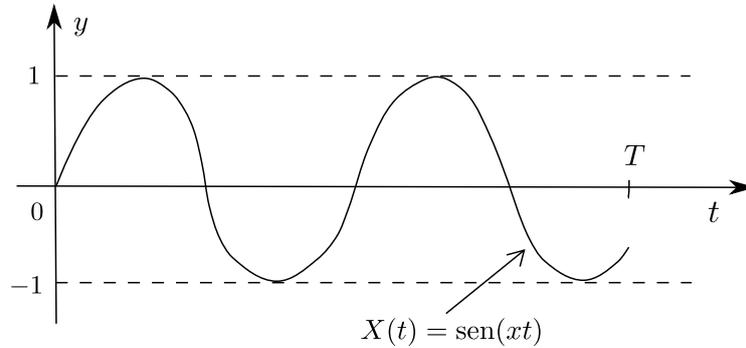


Figura 6: Gráfica de $\text{sen}(xt)$.

3. La recuperación del modelo de movimiento con velocidad aleatoria

Sea ahora T el tiempo (como en la Sección 1) de un viaje. De una manera bastante realista vamos a aceptar las siguientes hipótesis.

H₁. La velocidad $V(t, \omega)$ es un proceso *estocástico*.

H₂. $V(t, \omega)$ es un proceso *ergódico*.

En particular (Definición 2, inciso (a)), para cada $t \in [0, T]$,

$$\mathbb{E}V(t, \omega) = m_V = 60 \text{ (km/h)}.$$

Como la velocidad V depende del tiempo, la ecuación (4) ya no es aplicable, y debe ser reemplazada por la ecuación diferencial

$$ds = V(t, \omega) dt. \quad (14)$$

En un intervalo de tiempo infinitesimal dt la velocidad es constante. Integrando (14) y dividiendo entre T , obtenemos $\frac{1}{T} \int_0^S ds = \frac{1}{T} \int_0^T V(t, \omega) dt$, o bien (ya que $S = 120$),

$$\frac{120}{T} = \frac{1}{T} \int_0^T V(t, \omega) dt. \quad (15)$$

Ahora supongamos (no rigurosamente) que la distancia S es “relativamente grande” (como $S = 120$ km en nuestro ejemplo). Esto último implica que la v.a. T (el tiempo de viaje) es también “bastante grande” lo que nos permite utilizar la igualdad aproximada:

$$\frac{1}{T} \int_0^T V(t, \omega) dt \approx \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T V(t, \omega) dt = \mu_V. \quad (16)$$

El último límite existe como consecuencia de la Definición 2 (b). Comparando las igualdades (15), (16), y (13) se obtiene que $\frac{120}{T} \approx m_V = 60$, o bien $T \approx \frac{120}{60} = 2$. La última aproximación significa que todos los valores de la v.a. T están cercanos a la constante 2, y esto permite conjeturar que $\mathbb{E}T \approx 2$ (h), es decir, el tiempo promedio de viaje es alrededor de dos horas. De tal manera hemos recuperado nuestra conjetura intuitiva (1) (ver la Sección 1). Sin embargo, la afirmación $\mathbb{E}T \approx 2$ es cierta solamente para viajar grandes distancias.

4. Procesos Estocásticos Estacionarios en Sentido Amplio

Sea $X(t, \omega)$ un proceso estocástico ($t \in [0, \infty]$ o $t = 1, 2, \dots$). En las definiciones de procesos estacionarios vamos a utilizar la siguiente hipótesis:

H₃. Para cada t , la esperanza $\mathbb{E}X(t, \omega)$ y la varianza $\text{Var}(X(t, \omega))$ son finitas, es decir, se tiene

$$E|X(t, \omega)|^2 < \infty.$$

Denotaremos: $X_t \equiv X(t, \omega)$, $t \geq 0$.

Definición 3 La función de covarianza del proceso X_t se define como:

$$\rho(t, s) := \text{Cov}(X_t, X_s) := E[(X_t - EX_t)(X_s - EX_s)] = EX_t X_s - EX_t EX_s,$$

para $t, s \in [0, \infty]$ o $t, s \in \{0, 1, 2, \dots\}$.

Definición 4 Se dice que el proceso estocástico X_t es estacionario en sentido amplio si:

- (a) $EX_t = m_X$, o sea, no depende del tiempo.
- (b) La función de covarianza $\rho(t, s)$ depende únicamente del tamaño del incremento de tiempo, es decir, si denotamos por $\tau = |t - s|$, se tiene que:

$$\rho(t, s) = \rho(|t - s|) = \rho(\tau).$$

Ejemplo 3 (a) **Proceso Autorregresivo.** Sea $m \geq 1$ un entero dado y sean

$$\xi_{1-m}, \xi_{2-m}, \dots, \xi_0, \xi_1, \xi_2, \dots$$

v.a. independientes e idénticamente distribuidas, tales que $E\xi_k = 0$, $E|\xi_k|^2 < \infty$. Ahora definamos:

$$X_n = \xi_n + \alpha_1 \xi_{n-1} + \alpha_2 \xi_{n-2} + \cdots + \alpha_m \xi_{n-m}, \text{ donde } \alpha_i \in \mathbb{R}, i = 1, 2, \dots, m.$$

En el caso particular donde $m = 1$, tenemos que:

$$X_n = \xi_n + \alpha \xi_{n-1},$$

y se calcula fácilmente, (consultar, por ejemplo (5))

$$\rho(\tau) = \begin{cases} \text{Var}(\xi_0)(1 + \alpha^2) & \text{si } \tau = 0, \\ \alpha \text{Var}(\xi_0) & \text{si } \tau = 1, \\ 0 & \text{si } \tau > 1. \end{cases} \quad (17)$$

Entonces para $m = 1$, X_n es un proceso estacionario.

(b) **Incrementos del proceso de Poisson.** Sea $N(t)$, $t \geq 0$ un proceso de Poisson con parámetro $\lambda > 0$, así que $\mathbb{E}N(t) = \lambda t$. Definamos:

$$X(t) = N(t+1) - N(t), \quad t \geq 0.$$

El proceso $X(t)$ es estacionario puesto que primeramente $\mathbb{E}X(t) = \mathbb{E}(N(t+1) - N(t)) = \lambda(t+1) - \lambda t = \lambda$. Ahora, teniendo en cuenta que el proceso de Poisson tiene incrementos independientes, no es difícil probar que:

$$\rho(\tau) = \begin{cases} \lambda(1 - \tau) & \text{si } |\tau| < 1, \\ 0 & \text{si } \tau \geq 1. \end{cases} \quad (18)$$

5. Leyes de los Grandes Números y Ergodicidad

Empecemos con algo simple. Una moneda simétrica se lanza n veces. Definamos las siguientes variables aleatorias para $k = 1, 2, \dots, n$:

$$X_k = \begin{cases} 1 & \text{si en el } k\text{-ésimo lanzamiento sale águila} \\ 0 & \text{si en el } k\text{-ésimo lanzamiento sale sol.} \end{cases}$$

Sea $S_n = X_1 + X_2 + \cdots + X_n = \#$ de sumandos iguales a 1
= al número de veces que ha salido águila en los n lanzamientos.

Por lo tanto, $\frac{S_n}{n}$ es la “frecuencia” de veces que sale “águila”. Por simetría de la moneda e intuición, cuando n es “grande”, debe de resultar:

$$\frac{S_n}{n} = \frac{\# \text{ de águilas en los } n \text{ lanzamientos}}{n} \approx \frac{1}{2} = P(\text{sale “águila”}).$$

Notemos que:

$$\mathbb{E}X_k = 1 \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} = \frac{1}{2}.$$

Teorema 1 (Bernoulli, 1714) Cuando $n \rightarrow \infty$ la sucesión de v.a. S_n/n converge en probabilidad a $1/2 = \mathbb{E}X_1 = P(\text{sale "águila"})$.

El siguiente teorema es un resultado mucho más profundo y general.

Teorema 2 (Kolmogorov, 1932) Sean X_1, X_2, \dots v.a. independientes e idénticamente distribuidas, y $S_n = X_1 + X_2 + \dots + X_n$. Supongamos que la esperanza $m = \mathbb{E}X_1 (= \mathbb{E}X_k)$ es finita. Entonces, con probabilidad 1,

$$\frac{S_n}{n} \rightarrow m, \text{ cuando } n \rightarrow \infty.$$

Existen diferentes versiones del Teorema 2 para procesos estacionarios. Como ejemplo, mencionamos el siguiente resultado.

Teorema 3 Sea X_t un proceso estacionario tal que su función de covarianza satisface la siguiente condición: existe un $\varepsilon > 0$ tal que para toda τ suficientemente grande

$$|\rho(\tau)| \leq \frac{1}{\tau^\varepsilon}. \quad (19)$$

Entonces con probabilidad 1 se tiene:

(a) En tiempo discreto, cuando $n \rightarrow \infty$

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow m_X = \mathbb{E}X_0.$$

(b) En tiempo continuo, cuando $T \rightarrow \infty$,

$$\frac{1}{T} \int_0^T X_t dt \rightarrow m_X = \mathbb{E}X_0. \quad (20)$$

Mas detalles sobre el Teorema 3 se puede consultar en (3; 7).

Nota 1 (a) El Teorema 3, (b) afirma que si el proceso $X(t)$ es estacionario y se cumple la condición (19), entonces $X(t)$ es un proceso ergódico (véase la Definición 2).

(b) La condición (19) (que se cumple particularmente para los procesos dados en el Ejemplo 3, (véase (17) y (18)) expresa una propiedad que $X(t)$, en un sentido "olvida su pasado".

Sean X v.a. de cuadrado integrable y no constante, y $X_1 = X_2 = \dots = X$, entonces X_n , $n > 1$ es un proceso estacionario, sin embargo "no es ergódico" (¡no olvida su pasado!) y además:

$$\frac{X_1 + X_2 + \dots + X_n}{n} = \frac{nX}{n} = X \not\rightarrow EX.$$

Ejemplo 4 Sea X_n , $n = 1, 2, \dots$ un proceso estacionario, en tiempo discreto, donde X_n representa la ganancia, (cuando $X_n > 0$) o pérdida (cuando $X_n < 0$) de un jugador en la n -ésima partida de un juego de azar en un casino. En cualquier casino se tiene que $m_X = EX_1 < 0$ (juego injusto). Ahora sean x_0 el capital inicial del jugador y $Z_n = x_0 + X_1 + \dots + X_n$ la fortuna del jugador después de n repeticiones del juego.

Suponiendo (19) (¡las v.a. podrían ser dependientes!) y teniendo en cuenta que $x_0/n \rightarrow 0$, tenemos que **con probabilidad 1**,

$$\frac{Z_n}{n} = \frac{x_0}{n} + \frac{X_1 + X_2 + \dots + X_n}{n} \rightarrow m_X < 0, \text{ (cuando } n \rightarrow \infty \text{)}.$$

Así que para un n “suficientemente grande” (caso de un jugador insistente) $Z_n < 0$, lo que implica la ruina inevitable del jugador.

6. Ergodicidad de Procesos de Markov

Los *procesos de Markov* se utilizan profusamente para modelar sistemas estocásticos donde la evolución futura del proceso sólo depende del presente, o sea, fijando un estado presente, los estados pasados del proceso no afectan a las predicciones sobre la evolución futura del proceso.

Ejemplo 5 (El proceso de difusión) Sea X_t , $t \geq 0$ el proceso estocástico que representa la posición de una molécula de tinta en el instante t . (Esta molécula se desplaza a causa de los golpes de las moléculas de agua.)

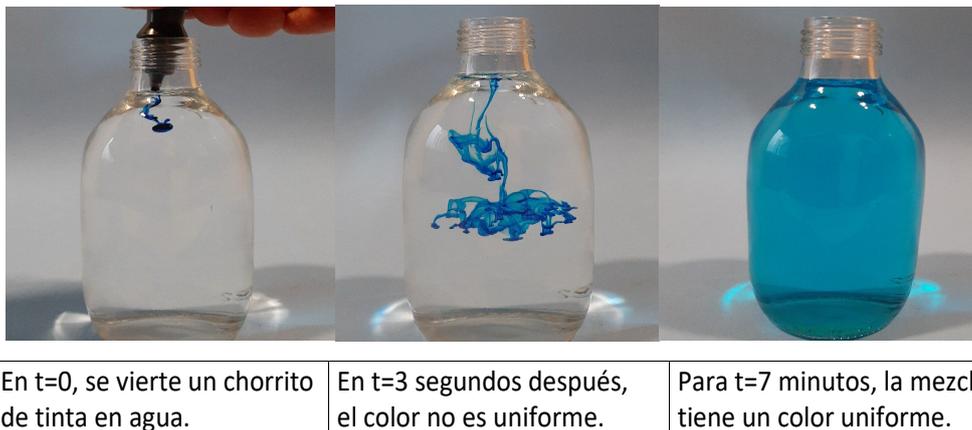


Figura 7: Esta figura muestra cómo se va haciendo la mezcla de la tinta con el agua.

Si en el instante t la posición de la molécula es $X_t = \bar{x}$, entonces en un tiempo futuro $t + \delta$ la molécula se ha desplazado, de manera aleatoria, a un conjunto A con una probabilidad dada por cierta función $p(\bar{x}, A)$. Esta función está definida para cada $\bar{x} \in L$ y $A \subset L$. Esta función se le llama la *probabilidad de transición* del proceso X_t . Lo que es importante aquí es el hecho de que la probabilidad $p(\bar{x}, A)$ (de moverse de \bar{x} al conjunto A) no depende de posiciones $X(s)$ ($s < t$) en donde la molécula haya estado antes del instante t . Es decir, $X(t)$, $t \geq 0$ es un proceso de Markov con los estados en el conjunto L de dimensión tres.

Es claro que X_t **no es un proceso estacionario**, sin embargo, **es ergódico** en el siguiente sentido:

- $X(t)$ “olvida su pasado” (sus valores pasados);
- conforme aumenta el tiempo t , las distribuciones de los vectores aleatorios de $X(t)$ “se estabilizan” y se aproximan a la distribución uniforme.

Para aclarar lo que se ha dicho, denotemos por $D(X(t))$ la distribución del vector aleatorio $X(t)$ y por U la distribución uniforme en el conjunto L . Se puede demostrar que:

- Cuando $t \rightarrow \infty$, $D(X(t)) \rightarrow U$ (conforme pasa el tiempo el color de la mezcla va siendo azul uniforme).
- Para t “grandes” el proceso $X(t)$ se aproxima a un proceso estacionario, por eso se comporta como un proceso estacionario.
- Con probabilidad 1,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T X(t) dt = \lim_{t \rightarrow \infty} \mathbb{E}X(t) = \bar{x}_*, \quad (21)$$

donde \bar{x}_* denota el vector de coordenadas del centro del recipiente.

Generalizando la Definición 2, se dice que un proceso de Markov $X(t)$ es *ergódico* si existe un estado \bar{x}_* tal que (21) se cumple con probabilidad 1 (para casi todos los estados iniciales).

Ejemplo 6 (La transmisión de información con errores) (ver (2)) Sea ξ_1, ξ_2, \dots una sucesión de v.a. independientes e idénticamente distribuidas tales que

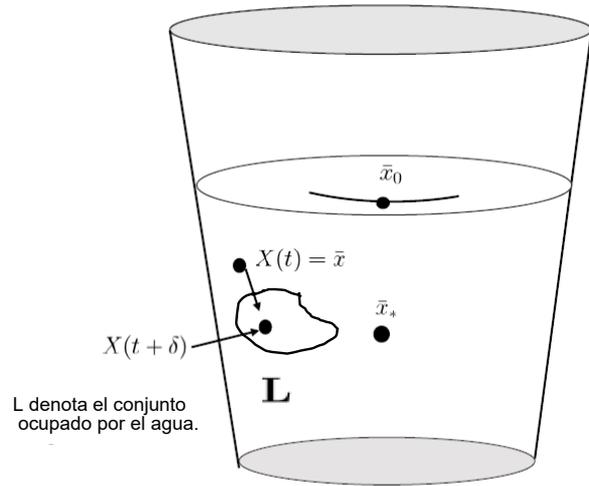


Figura 8: Esta figura muestra el desplazamiento a causa de los golpes de las moléculas de agua.

$$\xi_1 = \begin{cases} 1 & \text{con probabilidad } p \\ -1 & \text{con probabilidad } 1 - p, \end{cases} \quad (22)$$

donde $p \in (0, 1)$ es un parámetro dado.

Ahora utilizando estas v.a. se define el siguiente proceso estocástico a tiempo discreto: (empezando con $n = 0$).

$$X_0 \stackrel{\text{def}}{=} 1; \quad (23)$$

$$X_n = \prod_{i=1}^n \xi_i, \quad n = 1, 2, \dots \quad (24)$$

Es fácil probar que X_n es un proceso de Markov. Además, es claro que X_n puede tomar sólo dos valores; -1 y 1.

Calculemos las probabilidades

$$p_n = P(X_n = 1) \text{ y } 1 - p_n = P(X_n = -1).$$

Por un lado,

$$\mathbb{E}X_n = 1 \cdot p_n + (-1)(1 - p_n) = 2p_n - 1. \quad (25)$$

Por otro lado, aplicando (22) y (24) y tomando en cuenta la independencia de las v.a. ξ_1, ξ_2, \dots , obtenemos ($n = 1, 2, \dots$):

$$\mathbb{E}X_n = \prod_{i=1}^n \mathbb{E}\xi_i = \prod_{i=1}^n (2p - 1) = (2p - 1)^n. \quad (26)$$

Comparando (25) y (26) se obtiene que cuando $n \rightarrow \infty$,

$$p_n = P(X_n = 1) = \frac{(2p - 1)^n + 1}{2} \rightarrow \frac{1}{2}, \quad (27)$$

$$1 - p_n = P(X_n = -1) = \frac{1 - (2p - 1)^n}{2} \rightarrow \frac{1}{2}. \quad (28)$$

Tal convergencia de las distribuciones $D(X_n)$ a la distribución límite, uniforme $D_\infty = (1/2, 1/2)$ permite demostrar que el proceso de Markov X_n es ergódico.

El proceso X_n admite la siguiente interpretación. Supongamos que $X_0 = 1$ significa que alguien recibió una información verdadera. El/ella transmite esta información al receptor 1 correctamente con probabilidad p , e invierte el contenido del mensaje (es decir, manda un mensaje falso) con probabilidad $1 - p$. El receptor 1 retransmite a un receptor 2 la información que ha recibido, utilizando la misma regla:

- con probabilidad p envía sin cambios la información que ha recibido;

- con probabilidad $1-p$ el/ella altera el significado del mensaje.

El proceso de transmisión se repite (para el receptor 3, ..., receptor n).

Vamos a decir que el n -ésimo receptor recibe la información correcta si $X_n = 1$. En caso contrario, si $X_n = -1$ el receptor recibe la información falsa. Entonces, si por ejemplo $X_{n-1} = 1$, entonces de acuerdo a (24),

$$\begin{aligned} P(X_n = 1/X_{n-1} = 1) &= P(\xi_n = 1) = p, \\ P(X_n = -1/X_{n-1} = 1) &= P(\xi_n = -1) = 1 - p. \end{aligned}$$

La probabilidad del error $1 - p$ en la transmisión de la información (intencional o accidental) puede ser pequeña. Por ejemplo, sea $1 - p = 0.001$, es decir $p = 0.999$. Así de (27) se tiene que $P(X_2 = 1) \approx 0.998$, y también todos los n -ésimos receptores con n no muy grande va a recibir la información correcta con una alta probabilidad. Pero, por ejemplo, para $n = 5000$, $P(X_{5000} = 1) \approx 0.5000225$, y, correspondientemente, $P(X_{5000} = -1) = P(\text{el } 5000\text{-ésimo receptor va a recibir la información falsa}) \approx 0.4999775$.

Las relaciones límites en (27) y (28) indican que para n suficientemente grande el n -ésimo receptor van a recibir la información correcta o falsa con prácticamente las mismas probabilidades (una incertidumbre máxima). Teniendo en cuenta el “modelo” anterior, uno podría cuestionar la validez de algunos estudios históricos.

Agradecimientos

Los autores de este trabajo desean agradecer a los revisores por sus correcciones, comentarios y sugerencias que ayudaron a mejorar la presentación de este trabajo.

Referencias

- [1] A. Einstein, *Investigations on the theory of Brownian movement*. USA: Dover, 1956.
- [2] E. Gordienko and X. Popoca, *Introducción a la teoría de probabilidad y métricas probabilísticas con aplicaciones en seguros y finanzas*. México: Aportaciones Matemáticas, Instituto de Matemáticas UNAM, 2018.
- [3] T.-C. Hu, A. Rosalsky, and A. Volodin, "On Convergence properties of sums of dependent random variables under second moment and covariance restrictions," *Statistics and Probability Letters*, vol. 78, no. 14, pp. 1999–2005, 2008. DOI: <https://doi.org/10.1016/j.spl.2008.01.073>
- [4] J. Kahane, *Le mouvement brownien. Un essai sur les origines de la théorie mathématique*. Paris Francia: Société Mathématique de France, Séminaires et Congrès 3, 1998.
- [5] S. Meyn and R.L.Tweedie, *Markov Chains and Stochastic Stability*. New York: Cambridge University Press, 2009.
- [6] J. Perrin, "Mouvement brownien et réalité moleculaire," *Annales de chimie et de physique 8è série*, vol. 18, pp. 5–114, 1909.
- [7] A. Poznyak, "A new version of the strong law of large numbers for dependent vector processes with decreasing correlation," in *Proceedings of the 39th IEEE Conference on Decision and Control (Cat. No.00CH37187)*, vol. 3, 2000. DOI: 10.1109/CDC.2000.914247 pp. 2881–2882.
- [8] A. Shiryaev, *Probability*. New York: Springer-Verlang, 1996.

Como citar este artículo: E. Gordienko and J. Ruiz de Chávez , "Un viaje con velocidad aleatoria y procesos estocásticos ergódicos", Sahuarus. Revista Electrónica de Matemáticas, vol. 6, no. 1, pp 1-15, 2022. <https://doi.org/10.36788/sah.v6i1.126>

Border ideals: un algoritmo de pertenencia

Petra Rubí Pantaleón-Mondragón¹

¹UMSNH-CCM

¹pantaleon.rubi@gmail.com

Resumen

Dado un entero positivo N fijo, se considera el esquema de Hilbert de N puntos sobre el plano complejo, denotado por $\mathbb{H}^N(\mathbb{C}^2)$. Una partición λ de N define un abierto U_λ en $\mathbb{H}^N(\mathbb{C}^2)$. El propósito principal de este artículo es presentar un algoritmo para, dado un ideal 0-dimensional $I \subset \mathbb{C}[x, y]$, encontrar todas las particiones λ de N tales que $I \in U_\lambda$.

Palabras Clave: Órdenes Monomiales; Esquema de Hilbert; Particiones.

DOI: 10.36788/sah.v6i1.127

Recibido: 6 de junio de 2022.

Aceptado: 6 de septiembre de 2022.

1. Introducción

Uno de los objetos matemáticos interesantes e importantes actualmente son los esquemas de Hilbert. En un contexto general, fueron introducidos por Alexander Grothendieck en los *Éléments de géométrie algébrique*, o EGA; el cual, es un tratado de 8 partes sobre Geometría Algebraica publicado entre 1960 y 1967. Los esquemas de Hilbert más estudiados son los llamados esquemas de Hilbert de puntos. Particularmente, existe mucha bibliografía con respecto a los esquemas de Hilbert de N puntos sobre el plano complejo. En este escrito, lo denotaremos por $\mathbb{H}^N(\mathbb{C}^2)$ para cualquier entero positivo N .

A groso modo, para $N \in \mathbb{Z}_{>0}$ fijo, el esquema de Hilbert $\mathbb{H}^N(\mathbb{C}^2)$ parametriza ideales $I \subset \mathbb{C}[x, y]$ cuyo conjunto de soluciones son N puntos (posiblemente iguales). En otras palabras, como conjunto, $\mathbb{H}^N(\mathbb{C}^2)$ consiste de ideales cuya variedad asociada consta de a lo más N puntos en el plano complejo. Con más estructura, se sabe que $\mathbb{H}^N(\mathbb{C}^2)$ es una variedad algebraica conexa, suave, irreducible y de dimensión $2N$ [5, 9]. Otra propiedad de este esquema, y la cual es mi interés en este escrito, es que posee una cubierta por abiertos afines U_λ , es decir, conjuntos abiertos pero con la estructura de variedad afín [13, 10]. Estos conjuntos abiertos U_λ son parametrizados por particiones λ de N , lo que hace posible poder trabajar en el esquema de manera combinatoria y no solo algebraica o geoméricamente [13].

Una de las motivaciones de este trabajo es determinar en qué abiertos U_λ del esquema de Hilbert pertenece un ideal $I \in \mathbb{H}^N(\mathbb{C}^2)$ fijo; más aún, determinar aquellos ideales que tienen la misma función Hilbert. Conocer esto proporciona una estratificación de nuestro esquema y esta puede ser usada para estudiar otro tipo de objetos [2, 6, 12].

Una primera forma de determinar dicha pertenencia es por medio de los ideales iniciales con respecto a un orden monomial, los cuales son ideales monomiales. Sin embargo, no todos los ideales monomiales provienen de un orden monomial; como veremos en el Ejemplo (21). El algoritmo de pertenencia que presentamos nos ayuda a determinar todos los abiertos sin el uso de órdenes monomiales.

En general, cuando se trabaja con más variables y sin órdenes monomiales, los ideales monomiales asociados a un ideal fijo son llamados “border ideals” los cuales son un problema interesante por sí mismos. Sus aplicaciones pueden ser usadas en problemas de programación lineal, teoría de códigos, etc [11].

La estructura del escrito es la siguiente. En la Subsección 2.1 introduciremos brevemente una definición del esquema de Hilbert de puntos sobre el plano complejo. En la Subsección 2.2 daremos definiciones básicas y generales para dar la correspondencia entre los ideales y sus ideales iniciales con respecto a un orden monomial fijo. En la Subsección 2.3 daremos el concepto de partición y su relación con un ideal monomial. Además, definiremos los abiertos que cubren al esquema de Hilbert. Finalmente, en la Sección 3 presentaremos el algoritmo de pertenencia, demostraremos su efectividad y daremos un par de ejemplos implementado en el software Macaulay2 [7]. Este algoritmo puede ser modificado e implementado en otros softwares matemáticos como Singular [8].

2. Preliminares

Los conceptos y resultados presentados en esta sección son de manera general, salvo cuando hablemos del esquema de Hilbert y particiones. Las demostraciones y detalles pueden ser revisadas principalmente en [3, 14, 13].

2.1. Esquema de Hilbert de puntos

Consideremos el anillo de polinomios con coeficientes complejos $\mathbb{C}[x_1, \dots, x_m]$. Recordemos que un **ideal** I es un subconjunto de $\mathbb{C}[x_1, \dots, x_m]$ el cual, es un grupo aditivo y satisface que $I \cdot \mathbb{C}[x_1, \dots, x_m] \subset I$.

Uno puede verificar que, para cualquier conjunto finito de polinomios no ceros, $f_1, \dots, f_l \in \mathbb{C}[x_1, \dots, x_m]$, el conjunto

$$\langle f_1, \dots, f_l \rangle := \left\{ \sum_{i=1}^l h_i f_i \mid h_i \in \mathbb{C}[x_1, \dots, x_m] \right\}$$

es un ideal, y es llamado ideal generado por f_1, \dots, f_l . Si un ideal es generado por monomios, decimos que es un **ideal monomial**. En general, el teorema de las bases de Hilbert [3, Teorema 4, §2.5] garantiza que, todo ideal $I \subset \mathbb{C}[x_1, \dots, x_m]$ es **finitamente generado**, es decir, existen polinomios $f_1, \dots, f_l \in \mathbb{C}[x_1, \dots, x_m]$ tales que $I = \langle f_1, \dots, f_l \rangle$.

A cada ideal $I \subset \mathbb{C}[x_1, \dots, x_m]$ se le puede asociar un objeto geométrico $V(I)$ llamado variedad afín asociada a I , variedad afín o simplemente variedad. Este conjunto está definido

como $V(I) := \{(a_1, \dots, a_m) \in \mathbb{C}^m \mid f(a_1, \dots, a_m) = 0 \forall f \in I\}$. Aunque la correspondencia no es única, a partir del ideal podemos obtener información de la variedad. Por lo que es natural, buscar caracterizar ideales con propiedades específicas. Por ejemplo, algunos invariantes de la variedad, obtenidas del ideal, pueden ser dados a través de la siguiente función.

Definición 1. Sea $I \subset \mathbb{C}[x_1, \dots, x_m]$ un ideal. La *función de Hilbert* de I , denotada por HF_I , es la función sobre los enteros no negativos definida por $s \mapsto \dim_{\mathbb{C}} \frac{\mathbb{C}[x_1, \dots, x_m]_{\leq s}}{I_{\leq s}}$, donde $\mathbb{C}[x_1, \dots, x_m]_{\leq s}$ es el espacio vectorial que consta de todos los polinomios de grado a lo más s , e $I_{\leq s}$ es el subespacio vectorial de $\mathbb{C}[x_1, \dots, x_m]_{\leq s}$ con polinomios en I .

Se puede demostrar que para un $s \in \mathbb{Z}_{\geq 0}$ suficientemente grande, la función de Hilbert es un polinomio con coeficientes enteros, el cual denotaremos por $HP_I(t)$ y es llamado *polinomio de Hilbert* (ver [3, Proposición 3, §9.3]).

Supongamos que $HP_I(t) = c_d t^d + c_{d-1} t^{d-1} + \dots + c_0 \in \mathbb{Z}[t]$. Definimos la *dimensión* de la variedad $V(I)$ como el grado del polinomio de Hilbert, es decir, $\dim V(I) = d$. En este texto, diremos que un ideal I es d -dimensional, haciendo referencia a la dimensión de $V(I)$. Definimos el *grado* de $V(I)$ como el número $\frac{c_d}{d!}$. Análogamente, diremos “grado de I ” en lugar de grado de la variedad.

Observemos que, un ideal es 0-dimensional si su polinomio de Hilbert es un polinomio constante. Por tanto, si fijando este polinomio constante, se tiene la siguiente definición.

Definición 2. Para $N \in \mathbb{Z}_{>0}$ fijo. El *esquema de Hilbert de N puntos sobre \mathbb{C}^2* se define como

$$\mathbb{H}^N(\mathbb{C}^2) := \{I \subset \mathbb{C}[x_1, x_2] \text{ ideal} \mid HP_I(t) = N\}.$$

La estructura de variedad se puede adquirir a través de un encaje de Plücker, el cual no discutiremos aquí pues se requiere más teoría de geometría algebraica, y para los fines del artículo no se es necesario, sin embargo, para ver más detalles sobre la estructura y sus propiedades se puede revisar [13, Cap. 18].

2.2. Bases de Gröbner

Definición 3. Un *orden monomial* es una relación \succ sobre $\mathbb{Z}_{\geq 0}^m$ que satisface tres propiedades:

- \succ es total, es decir, cualquiera dos elementos son comparables.
- Para $\alpha, \beta, \gamma \in \mathbb{Z}_{\geq 0}^m$. Si $\alpha \succ \beta$, entonces $\alpha + \gamma \succ \beta + \gamma$.
- \succ es un buen orden, es decir, cualquier subconjunto de $\mathbb{Z}_{\geq 0}^m$ tiene un elemento mínimo.

Veamos algunos ejemplos. Sean $\alpha = (\alpha_1, \dots, \alpha_m)$, $\beta = (\beta_1, \dots, \beta_m) \in \mathbb{Z}_{\geq 0}^m$ y un vector $w \in \mathbb{R}_{\geq 0}^m$.

Ejemplo 4. Orden graduado inverso lexicográfico; $\succ_{GRevLex}$. Decimos que $\alpha \succ_{GRevLex} \beta$ si $|\alpha| := \sum_{i=1}^m \alpha_i > |\beta| := \sum_{i=1}^m \beta_i$, o $|\alpha| = |\beta|$ y la primera entrada de la derecha no cero de $\alpha - \beta$ es negativa.

En concreto: $(3, 1) \succ_{GRevLex} (2, 2)$. Pues $|(3, 1)| = 4 = |(2, 2)|$ y en la diferencia $(3, 1) - (2, 2) = (1, -1)$ la última entrada no cero es negativa.

Observaciones 5. Un orden monomial que compara los grados totales, es decir, $|\alpha| \geq |\beta|$, como en el ejemplo anterior, son llamados órdenes monomiales graduados.

Ejemplo 6. Orden por vector de peso w ; \succ_w . Sea \succ un orden monomial cualquiera, decimos que $\alpha \succ_w \beta$ si $w \cdot \alpha > w \cdot \beta$ o $w \cdot \alpha = w \cdot \beta$ y $\alpha \succ \beta$, donde \cdot es el producto punto usual.

Por ejemplo: Si $w = (0, 2)$, entonces $(2, 2) \succ_w (3, 1)$ para cualquier orden monomial \succ . Como primero comparamos con el producto punto entonces, $(0, 2) \cdot (2, 2) = 4 > 2 = (0, 2) \cdot (3, 1)$ sin importar en orden monomial \succ .

Se puede verificar que el orden $\succ_{GRevLex}$ es un orden monomial mientras que el orden por vector de peso solo es un orden parcial. Sin embargo, siempre es posible elegir un vector de peso w en el cual \succ_w sea orden monomial [14].

Dado que existe una correspondencia biyectiva entre el conjunto de monomios mónicos y los puntos en $\mathbb{Z}_{\geq 0}^m$, es decir, un monomio $x^\alpha := x_1^{\alpha_1} \cdots x_m^{\alpha_m} \in \mathbb{C}[x_1, \dots, x_m]$ corresponde al punto $\alpha = (\alpha_1, \dots, \alpha_m) \in \mathbb{Z}_{\geq 0}^m$ y viceversa; el orden monomial de $\mathbb{Z}_{\geq 0}^m$ induce un orden monomial entre los monomios mónicos de $\mathbb{C}[x_1, \dots, x_m]$ de la siguiente forma: para cualquier orden monomial \succ , decimos $x^\alpha \succ x^\beta$ si, y sólo si $\alpha \succ \beta$ donde x^α, x^β son monomios mónicos en $\mathbb{C}[x_1, \dots, x_m]$.

Una vez fijando un orden monomial, como el orden es total, uno puede comparar los monomios de un polinomio.

Definición 7. Sean $f = \sum a_\alpha x^\alpha \in \mathbb{C}[x_1, \dots, x_m]$ un polinomio no cero y \succ un orden monomial fijo.

- El *soporte* de f es el conjunto $Supp(f) := \{\alpha \in \mathbb{Z}_{\geq 0}^m : a_\alpha \neq 0\}$.
- El *multigrado* de f es definido por $multideg(f) := \max\{\alpha : \alpha \in Supp(f)\}$ (el máximo con respecto al orden \succ).
- El *término inicial* de f es el monomio dado por el *coeficiente inicial* $a_{multideg(f)}$ y el *monomio inicial* $x^{multideg(f)}$, es decir, $in_\succ(f) := a_{multideg(f)} x^{multideg(f)}$.
- Si $I \subset \mathbb{C}[x_1, \dots, x_m]$ es ideal no cero, el *ideal inicial* de I es el conjunto $in_\succ(I) := \langle in_\succ(f) : f \in I \rangle$.

Observaciones 8. Cuando el orden es monomial, el ideal inicial es monomial. Sin embargo, si consideramos órdenes parciales, como el orden por vector de peso, el término inicial de un polinomio podrá no ser monomio, en consecuencia, su “ideal inicial” no necesariamente es monomial, aunque sí puede ser definido.

Ejemplo 9. Sea $f(x_1, x_2) = 4x_1^2x_2 - 15x_2^3 - x_1x_2 + x_1 \in \mathbb{C}[x_1, x_2]$. Como el soporte de f es el conjunto $\{(2, 1), (0, 3), (1, 1), (1, 0)\}$, usando el orden monomial $\succ_{GRevLex}$ tenemos $(2, 1) \succ_{GRevLex} (0, 3) \succ_{GRevLex} (1, 1) \succ_{GRevLex} (1, 0)$, entonces $in_{\succ_{GRevLex}}(f) = 4x_1^2x_2$. Así, su ideal inicial $in_\succ(\langle f \rangle) = \langle in_{\succ_{GRevLex}}(f) \rangle$.

En el ejemplo anterior, el ideal inicial es el ideal monomial generado por el término inicial. Sin embargo, puede ocurrir que los generadores del ideal inicial se puedan obtener por cancelaciones de términos mayores, es decir, el conjunto de generadores del ideal I podría no ser suficiente para que sus términos iniciales generen al ideal inicial, como podemos ver el siguiente ejemplo.

Ejemplo 10. Sea $I = \langle x_1^3 - 2x_1x_2, x_1^2x_2 - 2x_2^2 + x_1 \rangle \subset \mathbb{C}[x_1, x_2]$ un ideal. Usando el orden monomial \succ_{GRex} tenemos: $in_{\succ_{GRex}}(x_1^3 - 2x_1x_2) = x_1^3$ y $in_{\succ_{GRex}}(x_1^2x_2 - 2x_2^2 + x_1) = x_1^2x_2$. Por otro lado, dado que I es ideal, $-x_2(x_1^3 - 2x_1x_2) + x_1(x_1^2x_2 - 2x_2^2 + x_1) = x_1^2 \in I$. Sin embargo, $x_1^2 \notin \langle x_1^3, x_1^2x_2 \rangle$.

La siguiente definición nos permite obtener la igualdad entre el ideal inicial de un ideal y el ideal generado por sus términos iniciales.

Definición 11. Sea \succ un orden monomial fijo. Una *base de Gröbner* de un ideal I con respecto a \succ es un subconjunto finito $\mathcal{G} = \{g_1, \dots, g_s\}$ de I cuyos términos iniciales generan al ideal inicial de I , es decir,

$$\langle in_{\succ}(g_1), \dots, in_{\succ}(g_s) \rangle = in_{\succ}(I).$$

Buchberger demostró que todo ideal distinto del ideal cero en $\mathbb{C}[x_1, \dots, x_m]$ tiene una base de Gröbner (y por lo tanto, al menos un ideal monomial asociado). Más aún, dio un algoritmo que calcula una base de Gröbner para un orden monomial fijo conteniendo al conjunto generador del ideal [1]. Actualmente existen varios softwares matemáticos que pueden calcular una base de Gröbner, aunque dependiendo del orden monomial estas pueden ser difíciles de calcular.

Como las bases de Gröbner dependen de un orden monomial, y podemos definir órdenes monomiales usando vectores de pesos, en un anillo de polinomios con más de una indeterminada, existen una infinidad de órdenes monomiales. Sin embargo, uno puede ver que, cada ideal tiene un número finito de ideales iniciales.

Teorema 12. Cada ideal $I \subset \mathbb{C}[x_1, \dots, x_m]$ tiene un número finito de ideales iniciales distintos.

Demostración. Ver [14, Teorema 1.2].

Para un ideal fijo, un objeto que nos ayuda a determinar todos sus ideales iniciales distintos es el abanico de Gröbner, el cual explicaremos brevemente a continuación.

Sean $g = \sum a_{\alpha}x^{\alpha} \in \mathbb{C}[x_1, \dots, x_m]$ y $\omega \in \mathbb{R}^m$. Análogo a la definición de término inicial, definimos $in_{\omega}(g)$ como la suma de todos los términos $a_{\alpha}x^{\alpha}$ de g tales que el producto punto usual $\omega \cdot \alpha$ es máximo (con respecto al producto punto de ω con los elementos del soporte de g). Para cualquier ideal $I \subset \mathbb{C}[x_1, \dots, x_m]$, definimos el ideal de formas $in_{\omega}(I) := \langle in_{\omega}(f) \mid f \in I \rangle$. De hecho, para cualquier orden monomial \succ y cualquier ideal I , existe un vector $\omega \in \mathbb{Z}_{\geq 0}^m$ tal que $in_{\omega}(I) = in_{\succ}(I)$ (ver [14, Proposición 1.11]).

Sean $\omega \in \mathbb{R}^m$ y \succ un orden monomial fijos. Sea \mathcal{G} una base de Gröbner de I con respecto al orden \succ_{ω} . Definimos el cono abierto poliedral convexo $C[\omega] := \{\omega' \in \mathbb{R}^m \mid in_{\omega'}(g) =$

$in_{\omega'}(g) \forall g \in \mathcal{G}$. El conjunto de la cerradura de los conos $C[\omega]$ para toda $\omega \in \mathbb{R}^m$ es llamado **abanico de Gröbner**. En [14, proposición 2.4] se prueba, en efecto, que es un abanico, es decir, un complejo poliedral que consiste de un número finito de conos. Por lo tanto, podemos elegir un orden monomial en cada uno de los interiores de los conos, y este, nos determinará un ideal inicial distinto por cada uno de ellos. Actualmente, existen softwares matemáticos que nos permiten calcularlo. En el Ejemplo (21), calculamos el abanico de Gröbner para un ideal I con el uso del software Singular.

Algunos resultados importantes entre un ideal y su ideal inicial (con respecto a algún orden monomial) son los siguientes.

Proposición 13. *Sean $I \subset \mathbb{C}[x_1, \dots, x_m]$ un ideal y \succ un orden graduado. El ideal inicial $in_{\succ}(I)$ tiene la misma función de Hilbert que I .*

Demostración. *Ver [3, Proposición 4, §9.3].*

Si $M \subset \mathbb{C}[x_1, \dots, x_m]$ es un ideal monomial. Los monomios que no pertenecen a M son llamados **monomios estándar** de M .

Teorema 14 (Macaulay). *Sea $M \subset \mathbb{C}[x_1, \dots, x_m]$ un ideal monomial. Los monomios estándar de M forman una base del \mathbb{C} -espacio vectorial $\mathbb{C}[x_1, \dots, x_m]/M$.*

Demostración. *Ver [4, Teorema 15.3].*

Observaciones 15. *Observemos que una consecuencia del Teorema de Macaulay sobre un ideal monomial $M \subset \mathbb{C}[x_1, x_2]$ tal que $x_1^{\alpha_1}, x_2^{\alpha_2} \in M$, para algunos $\alpha_1, \alpha_2 \in \mathbb{Z}_{>0}$, implica que M es 0-dimensional, pues si $s_0 = \alpha_1 + \alpha_2$, entonces $HF_M(s) = HF_M(s_0)$ para toda $s \geq s_0$. Aunque el entero s que define al polinomio de Hilbert puede ser menor que s_0 , con esto podemos concluir que el polinomio de Hilbert de M es constante. De hecho, un ideal monomial 0-dimensional en $\mathbb{C}[x_1, x_2]$ contine a $x_1^{\alpha_1}, x_2^{\alpha_2}$ para algunos $\alpha_1, \alpha_2 \in \mathbb{Z}_{>0}$, de lo contrario, supongamos que x_1^{α} no pertenece al ideal monomial para toda $\alpha \in \mathbb{Z}_{>0}$, entonces los puntos de la forma $(a, 0) \in \mathbb{C}^2$ son elementos de la variedad asociada al ideal. Por tanto, no es de dimensión 0.*

2.3. Particiones

Por comodidad, cambiaremos la notación del anillo de polinomios que vamos a usar a partir de esta subsección y del resto del artículo. Nos centraremos en el anillo de polinomios $\mathbb{C}[x, y]$.

Definición 16. *Sea $N \in \mathbb{Z}_{>0}$ fijo. Una **partición** de N , denotada por $\lambda \vdash N$, es una sucesión decreciente $\lambda := (\lambda_1, \dots, \lambda_s)$ de enteros positivos tal que $\sum_{i=1}^s \lambda_i = N$.*

Ejemplo 17. *En la Tabla 1, se muestran todas las particiones para $N = 1, \dots, 5$.*

N	$\lambda \vdash N$
1	(1)
2	(2), (1, 1)
3	(3), (1, 1, 1), (2, 1)
4	(4), (1, 1, 1, 1), (2, 2), (2, 1, 1), (3, 1)
5	(5), (1, 1, 1, 1, 1), (3, 2), (2, 2, 1), (2, 1, 1, 1), (3, 1, 1), (4, 1)

Tabla 1: Particiones

Un diagrama de Young es un objeto combinatorio que consiste en un arreglo de casillas por filas, cada fila contiene una cantidad menor o igual de casillas a la anterior. A cada partición se le puede asociar un diagrama de Young. Por ejemplo, supongamos que tenemos la partición $\lambda = (4, 3, 2, 2, 1, 1) \vdash 13$, entonces podemos asociarle el diagrama de Young de la Figura (1a). Estos diagramas son conocidos como [escalera asociada a la partición \$\lambda\$](#) .

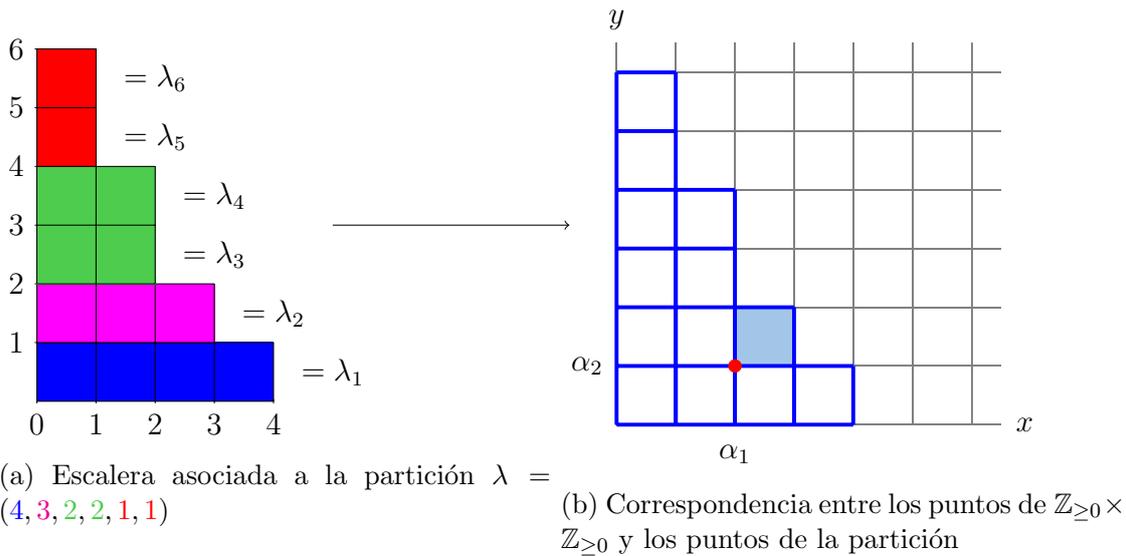


Figura 1: Partición, escalera en $\mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$

Por otro lado, una escalera la podemos pensar como una cuadrícula en $\mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$. Por tanto, a cada casilla se le puede asociar un punto $(\alpha_1, \alpha_2) \in \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$, el cual es el vértice inferior de la casilla, como podemos ver en el Figura (1b).

Más aún, por la correspondencia entre los puntos $(\alpha_1, \alpha_2) \in \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ y los monomios en $x^{\alpha_1}y^{\alpha_2} \in \mathbb{C}[x, y]$, podemos asociar a cada partición $\lambda = (\lambda_1, \dots, \lambda_s) \vdash N$ (o escalera) dos conjunto de polinomios en $\mathbb{C}[x, y]$:

- El conjunto de N monomios que corresponden a las casillas de la escalera asociada a λ , es decir,

$$B_\lambda := \{x^\delta y^\gamma : 0 \leq \delta \leq \lambda_i - 1, 0 \leq \gamma \leq i - 1, \text{ para } i = 1, \dots, s\}.$$



- El ideal generado por los monomios que corresponden al borde de la escalera

$$I_\lambda := \langle x^{\lambda_i} y^{i-1}, y^s : i = 1, \dots, s \rangle.$$

Este ideal es llamado el ideal asociado a la escalera (o partición) λ . En general, los elementos en el borde de la escalera no es un conjunto mínimo de generadores para I_λ . Por la observación (15), I_λ es un ideal monomial 0-dimensional en $\mathbb{C}[x, y]$. Además, podemos ver que B_λ es el conjunto de monomios estándar de I_λ .

Ejemplo 18. Consideremos la partición $\lambda = (4, 3, 2, 2, 1, 1) \vdash 13$, cuya escalera asociada es la dada en la Figura (1a). De la Figura (2) podemos ver que el conjunto de monomios estándar de I_λ es

$$B_\lambda = \{1, x, x^2, x^3, y, xy, x^2y, y^2, xy^2, y^3, xy^3, y^4, y^5\}.$$

Además, por definición, los generadores del ideal I_λ son

$$\{x^4, x^3y, x^2y^2, x^2y^3, xy^4, xy^5, y^6\}.$$

Sin embargo, observemos que los monomios de este conjunto que no están marcados en un círculo en la Figura (2), son múltiplos de los que sí lo están. Por tanto el ideal asociado a la partición está generado por los monomios $x^4, x^3y, x^2y^2, xy^4, y^6$, es decir, $I_\lambda = \langle x^4, x^3y, x^2y^2, xy^4, y^6 \rangle$.

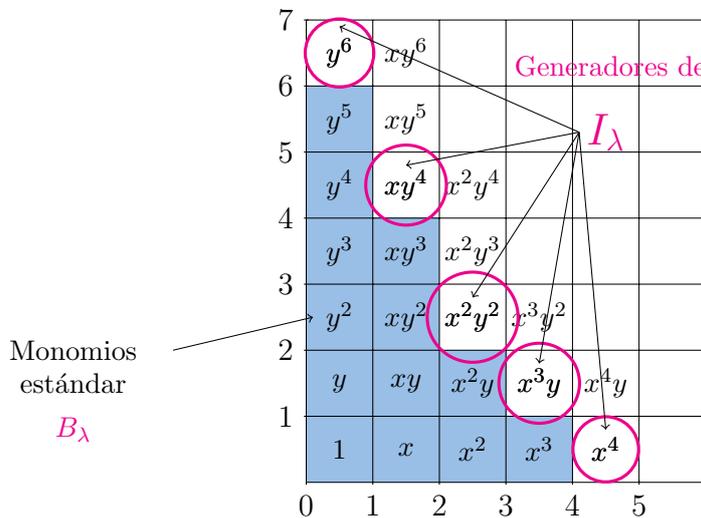


Figura 2: $(4, 3, 2, 2, 1, 1) \vdash 13$, I_λ , B_λ

Observaciones 19. A cada ideal monomial $M \in \mathbb{C}[x, y]$ 0-dimensional se le puede asociar una partición del grado de M . En efecto, si M es 0-dimensional, existen $\alpha_0, \alpha_1, \dots, \alpha_{s-1}, \beta_1, \dots, \beta_s \in \mathbb{Z}_{\geq 0}$ con $\alpha_0 > \alpha_1 > \dots > \alpha_{s-1}$ y $\beta_1 < \dots < \beta_s$ tales que

$$M = \langle x^{\alpha_0}, x^{\alpha_1} y^{\beta_1}, \dots, x^{\alpha_{s-1}} y^{\beta_{s-1}}, y^{\beta_s} \rangle.$$

Entonces, la partición asociada a M es

$$\lambda = (\underbrace{\alpha_0, \dots, \alpha_0}_{\beta_1}, \underbrace{\alpha_1, \dots, \alpha_1}_{\beta_2 - \beta_1}, \dots, \underbrace{\alpha_i, \dots, \alpha_i}_{\beta_{i+1} - \beta_i}, \dots, \underbrace{\alpha_{s-1}, \dots, \alpha_{s-1}}_{\beta_s - \beta_{s-1}}).$$

Además λ es un partición de $N = \sum_{i=1}^{s-1} \beta_i(\alpha_{i-1} - \alpha_i) + \beta_s \alpha_{s-1}$.

Definición 20. Para $N \in \mathbb{Z}_{>0}$ fijo y una partición λ de N . Sea $U_\lambda \subset \mathbb{H}^N(\mathbb{C}^2)$ el conjunto de ideales I tales que, bajo la proyección natural, el conjunto de monomios estándar B_λ forma una base del espacio vectorial $\frac{\mathbb{C}[x,y]}{I}$.

Observemos que los conjuntos $U_\lambda \neq \emptyset$ pues $I_\lambda \in U_\lambda$ para toda $\lambda \vdash N$. Por otro lado, sea $I \in \mathbb{H}^N(\mathbb{C}^2)$ y \succ un orden graduado, entonces, por la observación (19) y la Proposición (2.2), existe una partición $\lambda \vdash N$ tal que $in_\succ(I) \in U_\lambda$ y por lo tanto, $I \in U_\lambda$. Así, $\mathbb{H}^N(\mathbb{C}^2) = \bigcup_{\lambda \vdash N} U_\lambda$.

Para demostrar que los conjuntos U_λ son variedades abiertos afines, como mencionamos anteriormente, se utiliza el encaje de Plücker. Ellos se obtienen como los anulamientos de las coordenadas de este encaje (ver [13]).

Ejemplo 21. Consideremos el ideal $I \in \mathbb{H}^5(\mathbb{C}^2)$ generado por los siguientes 5 polinomios en $\mathbb{C}[x, y]$;

$$\begin{aligned} & x^3 - x, \\ & 2x^2y - y - y^2, \\ & 2xy - 2x - y + 2x^2 - y^2, \\ & 2xy^2 - 2x - y + 2x^2 - y^2, \\ & y^3 - y. \end{aligned}$$

Usando Singular, calculamos el abanico de Gröbner, y se obtuvo que I sólo tiene dos ideales iniciales diferentes obtenidas de las bases de Gröbner

$$\mathcal{G}_1 = \{y^3 - y, xy^2 - xy, x^2 + xy - \frac{1}{2}y^2 - x - \frac{1}{2}y\}$$

y

$$\mathcal{G}_2 = \{-2x^2 - 2xy + y^2 + 2x + y, x^2y - x^2 - xy + x, x^3 - x\}.$$

Por lo tanto, el ideal I pertenece a los abiertos $U_{\begin{smallmatrix} \square \\ \square \end{smallmatrix}}$ y $U_{\begin{smallmatrix} \square & \square \\ \square & \square \end{smallmatrix}}$. Por otro lado, si calculamos una base de Gröbner universal de I , es decir, la unión de las bases de Gröbner de I , obtenemos la siguiente base:

$$\mathcal{G} = \{x^2y - x^2 - xy + x, x^3 - x, y^3 - y, xy^2 - xy, x^2 + xy - \frac{1}{2}y^2 - x - \frac{1}{2}y\}$$



Para $\mathcal{G}_1 = \{y^3 - y, xy^2 - xy, x^2 + xy - \frac{1}{2}y^2 - x - \frac{1}{2}y\}$. El conjunto de términos iniciales de esta base es $\{y^3, xy^2, x^2\}$. Por lo tanto, la partición asociada es $(2, 2, 1)$.

4	y^3				
3	y^2	xy^2			
2	y	xy			
1	1	x	x^2		
	0	1	2	3	4

Figura 3: $\lambda = (2, 2, 1) \vdash 5$

Para $\mathcal{G}_2 = \{-2x^2 - 2xy + y^2 + 2x + y, x^2y - x^2 - xy + x, x^3 - x\}$. El conjunto de términos iniciales es $\{y^2, x^2y, x^3\}$. Por lo tanto, está asociado a la partición $(3, 2)$.

4					
3	y^2				
2	y	xy	x^2y		
1	1	x	x^2	x^3	
	0	1	2	3	4

Figura 4: $\lambda = (3, 2) \vdash 5$

4	y^3	xy^3			
3	y^2	xy^2			
2	y	xy	x^2y		
1	1	x	x^2	x^3	
	0	1	2	3	4

Figura 5: $\lambda = (3, 1, 1) \vdash 5$

Observemos que cada elemento de \mathcal{G} es una combinación de los monomios $\{1, x, x^2, y, y^2\}$ más términos del conjunto $\{x^3, xy, x^2y, xy^2, y^3\}$ que pertenecen al ideal $\langle x^3, xy, y^3 \rangle$, el cuál está asociado a la partición $(3, 1, 1)$ dada en la Figura 5. Sin embargo, no existe un orden monomial \succ tal que $in_{\succ}(I) = I_{\square\square}$.

En efecto, supongamos que existe un orden monomial \succ tal que $in_{\succ}(I) = \langle y^3, xy, x^3 \rangle$, entonces $in_{\succ}(x^2 + xy - \frac{1}{2}y^2 - x - \frac{1}{2}y) = xy$. Además $xy \succ x^2, y^2, x, y$. Como \succ es orden monomial, $xy \succ x^2$ implica que $y \succ x$, $yxy \succ y^2$ implica que $x \succ y$ así $x = y$, lo cual es una contradicción. Por lo tanto, el ideal I tiene asociado a los ideales monomiales $I_{\square\square}$, $I_{\square\square\square}$ y $I_{\square\square\square\square}$ y no todos son obtenidas a partir de un orden monomial.

Del ejemplo anterior surgen dos preguntas naturales ¿estas son todas las particiones de 5 para los cuales $I \in U_{\lambda}$? y si no, ¿cómo podemos determinar todas?

Para responder estas preguntas, diseñé un algoritmo que calcula (dado un ideal) todas las



particiones λ 's cuyos abiertos U_λ asociados contienen al ideal dado. Los cálculos consideran aquellas particiones que no provienen de órdenes monomiales.

3. Resultado

Recordemos que nuestro objetivo como parte de motivación para el caso general es considerar ideales $I \in \mathbb{C}[x, y]$ 0-dimensionales. Más aún, ideales $I \in \mathbb{H}^N(\mathbb{C}^2)$ con N fijo.

Algorithm 1 : Algoritmo de pertenencia

Entrada: $I \subset \mathbb{C}[x, y]$, ideal 0-dimensional.

Salida: Lista de todas las particiones λ del grado de I , tales que $I \in U_\lambda$.

- 1: Calcular una base ordenada $\mathcal{B} = \{x^{\alpha_1}y^{\alpha_2} + I\}$ del \mathbb{C} -espacio vectorial $\mathbb{C}[x, y]/I$.
 - 2: Calcular el conjunto de particiones Λ de $N = \dim_{\mathbb{C}} \mathbb{C}[x, y]/I$.
 - 3: Definir $P := \{\}$.
 - 4: Para cada $\lambda \in \Lambda$;
 - 5: Calcular el ideal monomial I_λ asociado a la partición λ .
 - 6: Calcular el conjunto de monomios estándar B_λ de I_λ .
 - 7: Calcular $\pi(B_\lambda)$, donde π es el morfismo proyección de $\mathbb{C}[x, y]$ a $\mathbb{C}[x, y]/I$.
 - 8: Definir la transformación $T_\lambda : \mathbb{C}[x, y]/I \rightarrow \mathbb{C}[x, y]/I$ que envía los elementos de la base ordenada \mathcal{B} a los elementos de $\pi(B_\lambda)$.
 - 9: Definir $r := \text{rank}(T_\lambda)$.
 - 10: Si $r = N$, entonces agregar λ a P .
 - 11: Regresa P .
-

Demostración (Prueba de validez). Sean $I \subset \mathbb{C}[x, y]$ un ideal 0-dimensional y \succ un orden monomial. Calculamos el ideal inicial de I con respecto al orden \succ . Como $\text{in}_\succ(I)$ es monomial, sea $\{x^{\alpha_1}y^{\alpha_2}\}$ el conjunto de monomios estándar de $\text{in}_\succ(I)$. Por el isomorfismo entre $\mathbb{C}[x, y]/\text{in}_\succ(I)$ y $\mathbb{C}[x, y]/I$ (ver [3]) entonces $\mathcal{B} := \{x^{\alpha_1}y^{\alpha_2} + I\}$ es una base del espacio $\mathbb{C}[x, y]/I$ y sin perder generalidad podemos suponer que es una base ordenada.

Sea $N = \dim_{\mathbb{C}} \mathbb{C}[x, y]/I$. Denotemos por Λ al conjunto de todas las particiones de N , es decir,

$$\Lambda := \left\{ \lambda = (\lambda_1, \dots, \lambda_s) \mid \sum_{i=1}^s \lambda_i = N, \text{ con } \lambda_1 \geq \dots \geq \lambda_s \geq 1 \right\}.$$

Sean $\lambda \in \Lambda$ y $B_\lambda := \{x^{\gamma_1}y^{\gamma_2}\}$ el conjunto de monomios estándar del ideal I_λ .

Consideremos el homomorfismo de anillos

$$\begin{aligned} \pi_\lambda : \mathbb{C}[x, y] &\rightarrow \mathbb{C}[x, y]/I \\ f &\mapsto f + I \end{aligned} .$$

Denotaremos por $\overline{B}_\lambda = \pi_\lambda(B_\lambda) \subset \mathbb{C}[x, y]/I$. Observemos que la cardinalidad $|\overline{B}_\lambda| = N$ (no necesariamente son independientes o diferentes de cero). Por tanto podemos definir la siguiente transformación lineal definida sobre la base \mathcal{B} :

$$\begin{aligned} T_\lambda : \mathbb{C}[x, y]/I &\rightarrow \mathbb{C}[x, y]/I \\ x^{\alpha_1}y^{\alpha_2} + I &\mapsto \pi_\lambda(x^{\gamma_1}y^{\gamma_2}) \end{aligned} .$$

Recordemos que T_λ es no singular si, y sólo si, T_λ es sobreyectiva. Por tanto, si $\text{rank}(T_\lambda) = N$, entonces T_λ es un isomorfismo, es decir, \overline{B}_λ es una base de $\mathbb{C}[x, y]/I$. Por tanto, por definición $I \in U_\lambda$.

Por otro lado, si $\text{rank}(T_\lambda) < N$, entonces \overline{B}_λ no es base de $\mathbb{C}[x, y]/I$, y por tanto, $I \notin U_\lambda$.

Ejemplo 22. Aplicaremos el Algoritmo (1) al ideal dado en el Ejemplo 21, es decir,

$$I = \langle x^3 - x, 2x^2y - y - y^2, 2xy - 2x - z + 2x^2 - z^2, 2xy^2 - 2x - y + 2x^2 - y^2, y^3 - y \rangle.$$

Usando el orden monomial $\succ_{GRevLex}$. Verificamos que el ideal I es 0-dimensional.

```
i1: R=QQ[x,y]
i2: I=ideal(x^3-x, 2x^2-2y-y^2, 2xy-2x-y+2x^2-y^2,
          2xy^2-2x-y+2x^2-y^2, y^3-y)
```

```
i3: dim I
o3: 0
```

Definimos el módulo M .

```
i4: M=R/I
```

Calculamos una base \mathcal{B} de M con respecto al orden $\succ_{GRevLex}$.

```
o5: {1, x, xy, y, y^2}
```

Definimos a N como el grado de I , o equivalentemente el número de elementos en la base $o5$.

```
i6: N=5
```

Calculamos el conjunto Λ de particiones de N .

```
o7: {(5), (4, 1), (3, 2), (3, 1, 1), (2, 2, 1), (2, 1, 1, 1),
      (1, 1, 1, 1, 1)}
```

Para cada partición, calculamos los bordes de las escaleras asociadas.



o8: $\{\{x^5, y\}, \{x^4, xy, y^2\}, \{x^3, x^2y, y^2\}, \{x^3, xy, xy^2, y^3\},$
 $\{x^2, x^2y, xy^2, y^3\}, \{x^2, xy, xy^2, xy^3, y^4\},$
 $\{x, xy, xy^2, xy^3, xy^4, y^5\}\}$

Para cada partición, calculamos los monomios estándar.

o9: $B = \{\{1, x, x^2, x^3, x^5\}, \{1, x, x^2, x^3, y\},$
 $\{1, x, x^2, y, xy\}, \{1, x, x^2, y, y^2\}, \{1, x, y, xy, y^2\},$
 $\{1, x, y, y^2, y^3\}, \{1, y, y^2, y^3, y^4\}\}$

Para cada elemento en cada lista de B , aplicamos el morfismo π .

o10: $\{\{1, x, -xy + 1/2y^2 + x + 1/2y, x, -xy + 1/2y^2 + x + 1/2y\},$
 $\{1, x, -xy + 1/2y^2 + x + 1/2y, x, y\},$
 $\{1, x, -xy + 1/2y^2 + x + 1/2y, y, xy\}, \{1, x, -xy + 1/2y^2 + x + 1/2y, y, y^2\},$
 $\{1, x, y, xy, y^2\}, \{1, x, y, y^2, y\},$
 $\{1, y, y^2, y, y^2\}\}$

Aplicamos la transformación T_λ .

o11:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & -1 & 0 & -1 \\ 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1/2 & 0 & 1/2 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 1 \\ 0 & 0 & 1/2 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1/2 & 1 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1/2 & 1 & 0 \\ 0 & 0 & 1/2 & 0 & 1 \end{pmatrix},$$

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

Calculamos el rango de cada transformación T_λ .

o12: $\{3, 4, 5, 5, 5, 4, 3\}$

Por último, elegimos las particiones cuya transformación T_λ tengan rango máximo.

o12: $\{(3, 2), (3, 1, 1), (2, 2, 1)\}$

Observemos que la lista obtenida al aplicar el algoritmo son justamente las tres particiones que obtuvimos en el Ejemplo 21.

Ejemplo 23. Consideremos el ideal $I = \langle -y^2 - x^2y, y + x^2 - xy^2 \rangle \in \mathbb{H}^7(\mathbb{C}^2)$. Existen 15 particiones de 7. Con el algoritmo de pertenencia, podemos ver que I pertenece a los abiertos U_λ , con $\lambda \in \{(7), (5, 2), (4, 3), (4, 2, 1), (4, 1, 1, 1), (3, 2, 1, 1, 1), (2, 2, 2, 1)\}$.

Agradecimientos

Muchas gracias a los dos revisores por sus sugerencias. El autor fue financiado por CONACyT.



Referencias

- [1] N. Bose, “Gröbner bases: An algorithmic method in polynomial ideal theory,” in *Multi-dimensional Systems Theory and Applications*. Springer, 1995, pp. 89–127.
- [2] J. Briançon, “Description de Hilb” $C\{x, y\}$,” *Inventiones mathematicae*, vol. 41, pp. 45–89, 1977.
- [3] D. Cox, J. Little, and D. OShea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*. Springer Science & Business Media, 2013.
- [4] D. Eisenbud, *Commutative algebra: with a view toward algebraic geometry*. Springer Science & Business Media, 2013, vol. 150.
- [5] J. Fogarty, “Algebraic families on an algebraic surface,” *American Journal of Mathematics*, vol. 90, no. 2, pp. 511–521, 1968. DOI: 10.2307/2373541
- [6] G. Gotzmann, “A stratification of the hilbert scheme of points in the projective plane,” *Mathematische Zeitschrift*, vol. 199, no. 4, pp. 539–547, 1988. DOI: 10.1007/BF01161642
- [7] D. R. Grayson and M. E. Stillman, “Macaulay2, a software system for research in algebraic geometry,” 2002.
- [8] G. Greuel, G. Pfister, H. Schönemann, O. Bachmann, W. Decker, C. Gorzel, H. Grassmann, A. Heydtmann, K. Krueger, M. Lamm *et al.*, “A computer algebra system for polynomial computations,” *Centre for Computer Algebra. University of Kaiserslautern*, 2005.
- [9] R. Hartshorne, “Connectedness of the hilbert scheme,” *Publications Mathématiques de l’IHÉS*, vol. 29, pp. 5–48, 1966. DOI: 10.1007/BF02684803
- [10] —, *Algebraic geometry*. Springer Science & Business Media, 2013, vol. 52.
- [11] A. Hashemi, M. Kreuzer, and S. Pourkhajouei, “Computing all border bases for ideals of points,” *Journal of Algebra and Its Applications*, vol. 18, no. 06, p. 1950102, 2019. DOI: 10.1142/S0219498819501020
- [12] A. Iarrobino, “Punctual hilbert schemes,” *Bulletin of the American mathematical society*, vol. 78, no. 5, pp. 819–823, 1972.
- [13] E. Miller and B. Sturmfels, *Combinatorial commutative algebra*. Springer Science & Business Media, 2005, vol. 227.
- [14] B. Sturmfels, *Grobner bases and convex polytopes*. American Mathematical Soc., 1996, vol. 8.

Como citar este artículo: P. R. Pantaleón-Mondragón. “Border ideals: un algoritmo de pertenencia.” *SAHUARUS. REVISTA ELECTRÓNICA DE MATEMÁTICAS*. ISSN: 2448-5365, 6(1), 16–29. <https://doi.org/10.36788/sah.v6i1.127>

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

José Batún-Cutz¹, Ruben Cool-Padilla², Ernesto Guerrero-Lara³, Diódora Kantún-Chim⁴ y Henry Panti-Trejo⁵

Facultad de Matemáticas, Universidad Autónoma de Yucatán.

¹jbatun@correo.uady.mx, ²ruben.cool@correo.uady.mx, ³ernesto.guerrero@correo.uady.mx, ⁴kchim@correo.uady.mx, ⁵henry.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 $S = \{1, 2, 3, 4, 5\}$.

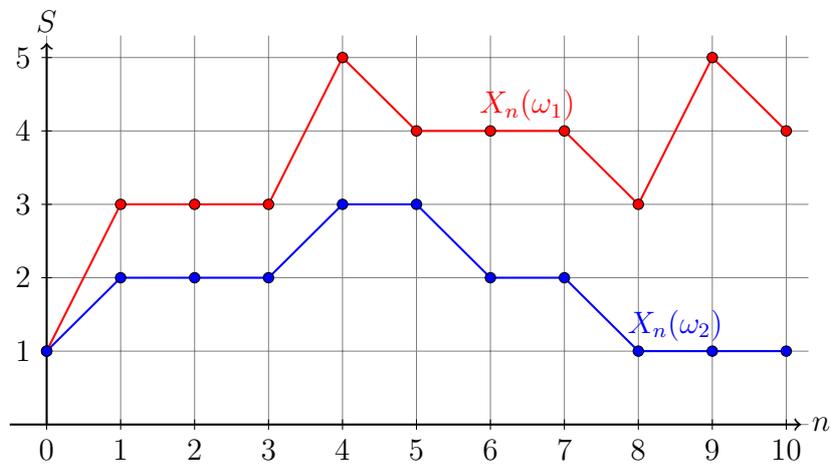


Figura 1: Ejemplo de trayectorias de un proceso estocástico $\{X_n, n \geq 0\}$

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.

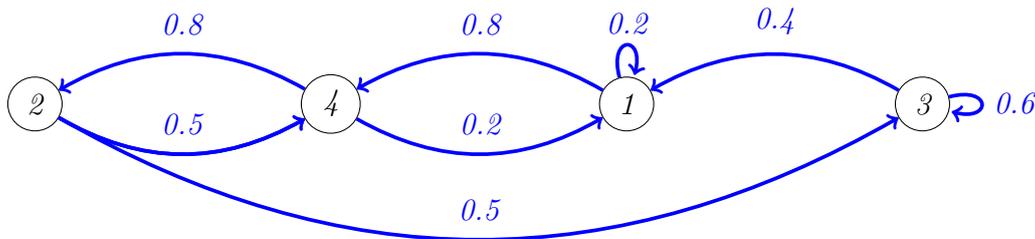
Una herramienta gráfica que permite visualizar de una forma más clara las transiciones entre estados en una cadena de Markov son los diagramas de transición, la cual es semejante a

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.

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

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

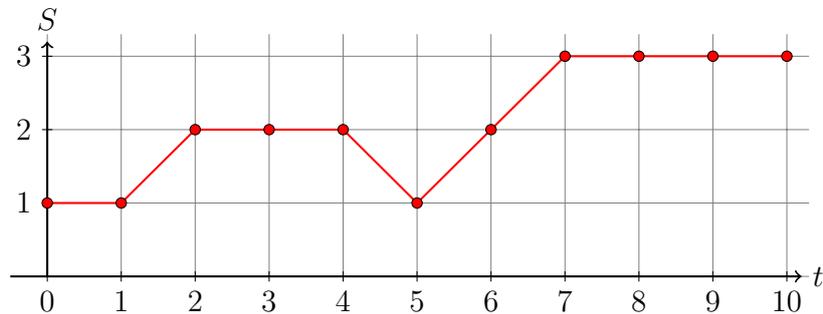


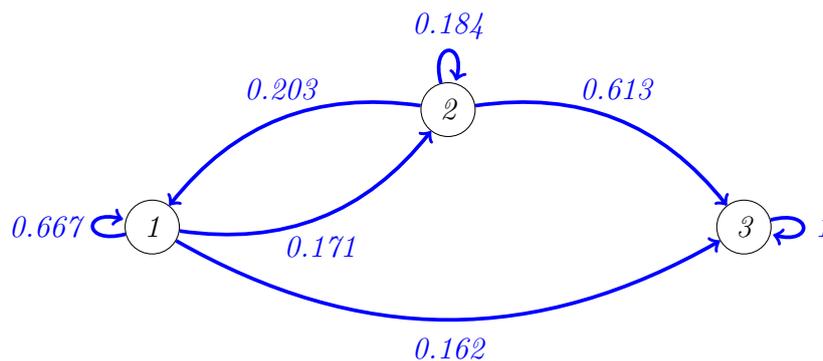
Figura 2: Trayectoria del paciente

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_α^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_α^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_α^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_α^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

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

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

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 λ :

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

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 λ está definida en la ecuación (16).

Región de rechazo:

Se rechaza H_0 si $U > \chi^2_{\rho, \alpha}$, 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>