



FACULTADE DE MATEMÁTICAS

Trabajo Fin de Grao

Principios básicos de las cadenas de Markov

Noelia Cebreiro Rodríguez

2020/2021

UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

GRADO DE MATEMÁTICAS

Trabajo Fin de Grao

Principios básicos de las cadenas de Markov

Noelia Cebreiro Rodríguez

Julio 2021

UNIVERSIDAD DE SANTIAGO DE COMPOSTELA

Trabajo propuesto

| |
|--|
| Área de Conocimiento: Estadística e investigación operativa |
| Título: Principios básicos de las cadenas de Markov |
| Breve descripción del contenido |
| En este trabajo se estudiará y presentará los resultados necesarios para entender los conceptos elementales de las cadenas de Markov |
| Recomendaciones |
| |
| Otras observaciones |
| |

Índice general

| | |
|--|------------|
| Resumen | VII |
| Introducción | 1 |
| 1. Fundamentos de probabilidad | 3 |
| 2. Procesos estocásticos | 9 |
| 3. Definiciones básicas de las cadenas de Markov | 13 |
| 4. Ecuaciones de Chapman-Kolmogorov | 19 |
| 5. Tiempos de primera pasada | 23 |
| 6. Clasificación de los estados | 29 |
| 7. Propiedades a largo plazo de las cadenas de Markov | 41 |
| 8. Simulación de cadenas de Markov en R | 47 |
| Bibliografía | 57 |

Resumen

El propósito de este trabajo es estudiar y entender las cadenas de Markov en tiempo discreto.

Comenzaremos con las nociones básicas de teoría de la probabilidad para llevar a cabo el posterior estudio de las cadenas de Markov.

Empezaremos por la definición de procesos estocásticos, así como la clasificación de estos y seguiremos con el estudio de las cadenas de Markov, hablaremos de sus propiedades así como de sus usos y aplicaciones más generales. Veremos, mediante ejemplos, la utilidad que tienen estas cadenas de Markov en la vida real.

Haremos especial hincapié en las cadenas de Markov homogéneas en tiempo discreto, nos centraremos en sus propiedades y daremos un resultado importante que nos indica cómo se comportan a largo plazo.

Por último, presentaremos algunos ejemplos de cadenas de Markov, utilizando la herramienta informática R, que nos permite hacer simulaciones, calcular características y obtener representaciones gráficas de las cadenas de Markov.

Abstract

The purpose of this work is to study and understand Markov's chains in discrete time. We will begin with the basic notions of probability theory to carry out the further study of Markov chains.

We will start with the definition of stochastic processes, as well as the classification of these and continue with the study of Markov chains. We will also refer to their properties as well as their more general uses and applications. We will see, through examples, how useful these Markov chains are in real life.

Special emphasis will be placed on homogeneous Markov chains in discrete time, focused on their properties and an important result that tells us how they behave in the long term

will be given.

Finally, some examples of Markov strings will be given, using the computer tool R, which allows us to make simulations, calculate characteristics and obtain graphical representations of Markov strings.

Introducción

La teoría de la probabilidad (ciencia matemática de la incertidumbre) juega un papel cada vez más importante en la forma que entendemos el mundo que nos rodea (clima, propagación de enfermedades, etc).

Por ejemplo, podríamos describir la evolución de un sistema a lo largo del tiempo mediante colecciones de variables aleatorias. De esta manera, se puede estudiar cómo evoluciona una variable aleatoria a lo largo del tiempo. Si tenemos la variable aleatoria “número de personas que esperan el metro en un instante t ”, entonces hay una variable aleatoria X_t para cada instante t en el que observamos el número de personas que esperan.

Así, la idea básica es identificar un proceso estocástico con una colección de variables aleatorias.

Las cadenas de Markov son un tipo concreto de estos procesos y son una herramienta esencial en el ámbito de la estadística y la investigación operativa. Nos permiten modelizar muchas situaciones de la vida real y de este modo poder estudiar con detalle procesos que ocurren a nuestro alrededor cada día, como podría ser predecir el tiempo de mañana, según el tiempo que ha hecho hoy.

Estaremos interesados en situaciones en las que pueden ocurrir una serie de eventos y donde la probabilidad de que ocurra un evento en el futuro, solo depende de lo ocurrido en el presente y es lo que distingue a las cadenas de Markov de otros procesos estocásticos.

Para hacernos una idea de todo esto podemos ilustrar con un ejemplo lo útil que puede llegar a ser una cadena de Markov en la vida cotidiana. En una empresa, se podría estudiar la preferencia de los clientes por tres marcas distintas de un producto, obteniendo mediante este estudio, los resultados de las preferencias en la actualidad. A partir de estos datos podríamos conocer cuáles serán las preferencias de los clientes por cada marca dentro de dos meses usando una cadena de Markov y así la empresa podrá elegir de una forma más útil en que marca invertir más y en cual menos.

Precedentes

Andréi Andréyevich Markov (1856-1922) fue un matemático ruso conocido por sus trabajos en la teoría de los números, así como en la teoría de probabilidades. Fue discípulo de Pafnuti Lvóvich Chebyshev en la Universidad de San Petersburgo , donde también impartió clases.

Markov influyó en diversos campos de las matemáticas pero su aportación más conocida fue la investigación en los procesos estocásticos que dieron lugar a las cadenas de Markov, las cuales fueron ideadas por el matemático ruso al estudiar la alternancia de vocales y consonantes en el poema Onegin de Pushkin.

Para ello desarrolló un modelo de probabilidad en el que los resultados de los sucesivos ensayos solo dependen de su predecesor inmediato y de ningún otro (esto lo consiguió extendiendo los resultados clásicos de sucesos independientes a cierto tipo de sucesos encadenados), de donde Markov obtuvo una descripción excelente de la alternancia de vocales y consonantes y le permitió calcular una estimación muy precisa de la frecuencia con la que ocurren las consonantes en el poema de Pushkin.

Un proceso de Markov permite modelizar la incertidumbre en muchos sistemas del mundo real que evolucionan dinámicamente en el tiempo.

Así, las cadenas de Markov han tenido gran utilidad en la economía, ingeniería, biología, etc.

También se han usado en las ciencias sociales, y particularmente en la lingüística y la hermenéutica (técnica o método de interpretación de textos) y son útiles en los cálculos probabilísticos meteorológicos y en muchas otras disciplinas.

Objetivos

El objetivo de este trabajo es hacer un análisis general de las cadenas de Markov, así como exponer sus utilidades en diversas ciencias y en la vida real.

Empezaremos con los fundamentos de probabilidad que serán necesarios en el estudio de las cadenas de Markov. Daremos la definición de proceso estocástico para después centrarnos en las cadenas de Markov caracterizadas por la propiedad markoviana, y seguiremos con el estudio de las características de las cadenas de Markov, ilustrándolo con ejemplos para su mejor conocimiento.

Capítulo 1

Fundamentos de probabilidad

Antes de comenzar, vamos a introducir los conceptos básicos de probabilidad que utilizaremos más adelante.

Cuando de un experimento podemos conocer cuál va a ser su resultado antes de que se realice, decimos que el experimento es determinista (por ejemplo las horas de salida del sol). Por el contrario, existen otros experimentos que no son deterministas (como conocer el número premiado de la lotería antes del sorteo).

Nosotros queremos estudiar experimentos que no son deterministas, pero no estamos interesados en todos ellos. Por ejemplo, no podremos estudiar un experimento del que no sabemos por anticipado los resultados que puede dar. Estudiaremos aquellos experimentos que no son deterministas y verifiquen ciertas condiciones que nos permiten un estudio riguroso de los mismos.

Llamaremos **experimento aleatorio** al que satisface los siguientes requisitos:

- Todos sus posibles resultados son conocidos de antemano.
- El resultado particular de cada realización del experimento es imprevisible.
- El experimento se puede repetir indefinidamente en condiciones idénticas.
- Se pueden asignar probabilidades a los resultados del experimento.

Cuando realizamos un experimento aleatorio, al conjunto de todos los posibles resultados lo llamamos **espacio muestral** (Ω).

Por ejemplo, si el experimento consiste en lanzar dos monedas, el espacio muestral es el conjunto $\Omega = \{(cara, cara), (cara, cruz), (cruz, cara) \text{ y } (cruz, cruz)\}$.

A cada uno de los elementos del espacio lo llamamos **suceso elemental**. Si consideramos el experimento aleatorio "lanzar una moneda", tendríamos el espacio muestral

$\Omega = \{\textit{cara}, \textit{cruz}\}$, donde un suceso elemental sería "que salga cara".

Y a cualquier subconjunto del espacio muestral lo llamaremos **suceso (o evento)**.

A continuación vamos a dar otras definiciones que usaremos más adelante:

Un **suceso seguro** es el suceso que siempre ocurre al realizar el experimento. En el ejemplo anterior, donde $\Omega = \{1, 2, 3, 4, 5, 6\}$, el suceso seguro sería "que salga cualquiera de los 6 resultados".

Un **suceso imposible**, al contrario que el anterior, y como indica su nombre, es aquel suceso que nunca ocurre y que denotamos por el vacío: \emptyset . Usando otra vez el experimento de lanzar un dado con espacio muestral $\Omega = \{1, 2, 3, 4, 5, 6\}$, tendríamos que el suceso imposible sería por ejemplo "sacar un número mayor que 8".

Lo que nos interesa es realizar operaciones con los sucesos y para ello necesitamos una estructura sobre el espacio muestral, de manera que cualquier operación que hagamos en el conjunto esté en dicha estructura.

Para ello definimos una **σ -álgebra** sobre un conjunto X como una familia \mathcal{A} no vacía de subconjuntos de X , cerrada bajo complementos, uniones e intersecciones numerables.

Los elementos de una σ -álgebra son los subconjuntos del espacio muestral. Una σ -álgebra es una estructura algebraica que se construye sobre Ω . Formalmente decimos que:

Una familia de subconjuntos de X , representada por \mathcal{A} , es una σ -álgebra sobre X cuando se cumplen las siguientes propiedades:

1. El conjunto vacío está en \mathcal{A} : $\emptyset \in \mathcal{A}$.
2. Si E está en \mathcal{A} , también está su complementario $E^c = X \setminus E$.
3. Si E_1, E_2, E_3, \dots es una sucesión de elementos de \mathcal{A} , entonces la unión (numerable) de todos ellos también está en \mathcal{A} .

Y ahora ya estamos en condiciones de definir la probabilidad:

La **probabilidad** es una función \mathbb{P} que asigna a cada elemento en la σ -álgebra de un espacio muestral, un número en el intervalo $[0, 1]$. Sea Ω un espacio muestral y \mathcal{A} una σ -álgebra sobre Ω . Decimos que $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ es una probabilidad en el espacio muestral Ω si verifica los siguientes axiomas:

- La probabilidad del suceso seguro es 1, es decir, $\mathbb{P}(\Omega) = 1$
- La probabilidad de la unión de sucesos es igual a la suma de las probabilidades de cada uno de esos sucesos, siempre que se trate de sucesos mutuamente excluyentes, es decir:

$$\mathbb{P}\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n \mathbb{P}[E_i] = \mathbb{P}[E_1] + \mathbb{P}[E_2] + \cdots + \mathbb{P}[E_n]$$

para cualquier $E_i \in \mathcal{A}$, con $i = 1, \dots, n$ tales que: $E_i \cap E_j = \emptyset$ para todo $i \neq j$.

Llamaremos **espacio de probabilidad** asociado a un experimento aleatorio a la terna $(\Omega, \mathcal{A}, \mathbb{P})$, donde Ω es el espacio muestral, \mathcal{A} es una σ -álgebra sobre Ω y por último, \mathbb{P} es una probabilidad definida sobre \mathcal{A} .

Supongamos ahora que en el estudio de un experimento aleatorio nos interesa conocer la probabilidad de que ocurra un cierto suceso A . Pero puede ser que dispongamos de información previa sobre el experimento una vez que sabemos que ha ocurrido el suceso B . Ahora la probabilidad de A puede que no sea la misma que cuando no sabíamos nada sobre B , es lo que llamamos **probabilidad condicionada**.

Formalmente decimos que dado un espacio de probabilidad $(\Omega, \mathcal{A}, \mathbb{P})$ y dos eventos (o sucesos) $A, B \in \mathcal{A}$ con $\mathbb{P}(B) > 0$, la **probabilidad condicionada** de A dado B está definida como:

$$\begin{aligned} \mathbb{P}(\cdot | B) : \mathcal{A} &\rightarrow [0, 1] \\ A &\rightarrow \mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \end{aligned}$$

Es decir, la probabilidad condicionada es la probabilidad de que ocurra un evento A , sabiendo que ha sucedido otro evento B .

Esta probabilidad es útil ya que como hemos dicho al principio, conocer un suceso que ha ocurrido nos fuerza a veces a "reajustar" la probabilidad de otro suceso como por ejemplo si definimos los siguientes sucesos al lanzar un dado:

A: "que salga el 3" $\implies \mathbb{P}(A) = \frac{1}{6}$,

B: "que salga impar" $\implies \mathbb{P}(B) = \frac{1}{2}$.

Si sabemos que ha sucedido B y queremos calcular la probabilidad de que suceda A tenemos una probabilidad condicionada y así $\mathbb{P}(A | B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$.

Así, si lanzamos un dado, la probabilidad de que salga 1 es $\frac{1}{6}$, pero si disponemos de la información adicional de que el resultado es impar reducimos los casos posibles de 6 a 3 (solo puede ser un 1, un 3 o un 5), con lo cual la probabilidad es $\frac{1}{3}$.

Al realizar un experimento aleatorio podemos estar interesados en asignar valores numéricos a los resultados. De manera informal, esta cantidad de interés se llama **variable**

aleatoria (variable porque toma distintos valores y aleatoria porque el valor observado no puede ser predicho antes de la realización del experimento).

Así una variable aleatoria es una función que asigna valores reales a los resultados de un experimento aleatorio.

Además una variable aleatoria puede ser **discreta**, si toma un número finito o infinito numerable de valores (x_1, \dots, x_i, \dots) o **continua** si toma valores en un intervalo.

Formalmente decimos que una variable aleatoria (v.a.) X es una función real

$$\begin{aligned} X : \Omega &\rightarrow \mathbb{R} \\ \omega &\rightarrow X(\omega) = x \end{aligned}$$

definida en el espacio de probabilidad en el espacio muestral Ω , asociado a un experimento aleatorio y que toma valores en \mathbb{R} .

Así para cada resultado del experimento aleatorio, $\omega \in \Omega$, obtenemos la medida numérica $X(\omega) \in \mathbb{R}$.

Pero no solamente estamos interesados en conocer los posibles valores de cualquier variable aleatoria, sino que también queremos conocer sus probabilidades de aparición mediante una función que asigna, a cada suceso definido sobre la variable, la probabilidad de que dicho suceso ocurra, es decir, nos interesan sus **distribuciones de probabilidad**. Pero para poder conocer estas distribuciones de probabilidad es necesario distinguir los distintos tipos de variables aleatorias.

Si tenemos una variable aleatoria discreta usaremos la **función de masa o de probabilidad** de una variable aleatoria discreta, X , que definimos como la función que asigna a cada valor su probabilidad, $p_i = \mathbb{P}(X = x_i)$.

Debemos tener en cuenta que si sumamos estas probabilidades sobre todos los posibles valores de X , la suma deberá ser 1: $\sum_i p_i = 1$.

Observemos que, conociendo la función de masa de una variable aleatoria discreta, se puede determinar la probabilidad de cualquier suceso relativo a la variable.

Por ejemplo, la probabilidad de que salga alguna cara al lanzar dos monedas, se puede identificar como la probabilidad de que la variable X =“número de caras” valga 1 ó 2. Sumando las probabilidades correspondientes resulta:

$$\mathbb{P}(\text{“que salga alguna cara”}) = \mathbb{P}(X \in \{1, 2\}) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = 0'5 + 0'25 = 0'75$$

Mientras que en el caso de una variable aleatoria continua, lo que tenemos es la **función de densidad** que es una aplicación $f : \mathbb{R} \rightarrow \mathbb{R}$ no negativa y tal que

$$\mathbb{P}(X \leq x_0) = \int_{-\infty}^{x_0} f(x)dx \quad \forall x_0 \in \mathbb{R}$$

Y por último, tanto para el caso discreto como para el caso continuo tenemos la **función de distribución** de una variable aleatoria, que definimos como:

$$\begin{aligned} F : \mathbb{R} &\rightarrow [0, 1] \\ x_0 &\rightarrow F(x_0) = \mathbb{P}(X \leq x_0) \end{aligned}$$

En este caso, lo único que cambia del caso discreto al caso continuo es su cálculo, así: Para el caso de una variable aleatoria discreta se tiene que

$$F(x_0) = \mathbb{P}(X \leq x_0) = \sum_{x_i \leq x_0} p_i$$

y esto se reduce a sumar las probabilidades de los valores menores o iguales que x_0 . Si la variable aleatoria es continua entonces la relación que se tiene es la siguiente:

$$F(x_0) = \mathbb{P}(X \leq x_0) = \int_{-\infty}^{x_0} f(x)dx.$$

Para terminar esta primera parte presentamos una medida de las variables aleatorias que utilizaremos muy a menudo y es lo que se conoce como la **esperanza** de una variable aleatoria y que representaremos por $\mathbb{E}[X]$.

Es el concepto análogo a la media aritmética cuando tenemos un conjunto de resultados de un experimento ya realizado. Para variables aleatorias se emplea el término esperanza, dando idea de que tenemos una variable aleatoria y si hay que aventurar un valor “esperado” de ella, ése sería el de la media (o esperanza). Formalmente decimos que:

- *Caso discreto:* Para una variable aleatoria discreta X con función de probabilidad $\mathbb{P}[X = x_i]$ con $i = 1, 2, \dots, n$ la esperanza se define como $\mathbb{E}[X] = \sum_{i=1}^n x_i \mathbb{P}[X = x_i]$
- *Caso continuo:* Para una variable aleatoria continua X con función de densidad $f(x)$ la esperanza se define como $\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)dx$

Con esto terminamos esta primera parte donde hemos introducido los conceptos básicos que utilizaremos a partir de ahora.

Capítulo 2

Procesos estocásticos

Como indicamos en la introducción, de manera informal decimos que los procesos estocásticos son modelos matemáticos que nos permiten representar muchos fenómenos que ocurren en la vida real.

Los procesos estocásticos tienen aplicaciones en múltiples campos y pueden ser útiles en cualquier momento en que se reconozca el papel de la aleatoriedad, la imprevisibilidad de eventos que pueden ocurrir en distintos momentos.

Veamos ahora la definición formal:

Definición 2.1. Un **proceso estocástico** es una colección de variables aleatorias $\{X_t/t \in T\}$, siendo

$$\begin{aligned} X_t : \Omega &\rightarrow \mathbb{R} \\ \omega &\rightarrow X_t(\omega) \end{aligned}$$

Donde T es el conjunto de índices del proceso estocástico (que identificaremos en muchas ocasiones con el tiempo).

Llamaremos **estado** a cada uno de los valores que pueden tomar las variables aleatorias. Y denotaremos por E al **espacio de estados** que definimos como el conjunto de posibles valores de las variables aleatorias X_t .

Clasificación de los procesos estocásticos:

Podemos tener un espacio de estados (E) discreto ó continuo. Por otro lado, el conjunto de índices puede ser de tipo discreto o de tipo continuo.

Y hay que tener en cuenta que todas las variables del proceso son del mismo tipo, es decir, todas las variables toman valores en el mismo espacio de estados (todas son discretas o todas son continuas).

Por tanto, dependiendo de cómo sea el conjunto de índices T y del espacio de estados E,

se puede establecer la siguiente clasificación de los procesos estocásticos:

- Si T es continuo, diremos que X_t es un proceso estocástico en tiempo continuo.
- Si por el contrario T es discreto, diremos que nos encontramos frente a un proceso estocástico en tiempo discreto.
- Si para cada instante t la variable aleatoria X_t es de tipo continuo, diremos que el proceso estocástico tiene espacio de estados continuo.
- Si para cada instante t la variable aleatoria X_t es de tipo discreto, diremos que el proceso estocástico tiene espacio de estados discreto.

| | T Discreto | T Continuo |
|-------------------|---|---|
| E Discreto | Proceso de estado discreto y tiempo discreto. | Proceso de estado discreto y tiempo continuo. |
| E Continuo | Proceso de estado continuo y tiempo discreto. | Proceso de estado continuo y tiempo continuo. |

A continuación vamos a ilustrar con un ejemplo lo que podría ser un proceso estocástico:

Ejemplo 2.2. El juego del monopolí se puede modelizar como un proceso estocástico donde X_k es la posición en el tablero del jugador después de k jugadas.

El espacio de estados es $\{1, 2, \dots, 40\}$ y denota las 40 casillas del tablero. Mientras que el conjunto de índices es $\{0, 1, 2, \dots\}$. Éstos dos conjuntos son discretos.

Una vez que hemos modelizado el juego mediante una cadena de Markov (que veremos más adelante) podríamos, por ejemplo, ver en qué casilla es más probable que caiga el jugador.

Ejemplo 2.3. Adaptaremos un ejemplo de Hillier para ilustrar diferentes conceptos y resultados que veremos a lo largo del trabajo.

Podemos modelizar el tiempo meteorológico de Ferrol mediante un proceso estocástico.

Supongamos que hay más posibilidades de que no llueva mañana si hoy no llueve. En particular la probabilidad de que mañana no llueva es de 0'8 si hoy no llueve y es de 0'6 si hoy llueve, y estas probabilidades no cambian si se considera la información del tiempo en los días anteriores.

Entonces, podemos ver como va a evolucionar el tiempo en Ferrol mediante un proceso

estocástico en tiempo discreto (días) con espacio de estados discreto $E = \{0, 1\}$.

Si comenzamos en un día inicial (etiquetado como día 0) y observamos el tiempo para cada día n , donde $n = 0, 1, 2, \dots$

El estado del sistema en el día n puede ser :

-Estado 0 = "no llueve".

-Estado 1 = "llueve".

Así para $n = 0, 1, 2, \dots$, la variable aleatoria X_n toma los valores:

$$X_n = \begin{cases} 0 & \text{si el día } t \text{ es seco.} \\ 1 & \text{si el día } t \text{ es lluvioso.} \end{cases}$$

El proceso estocástico $\{X_n, n = 0, 1, 2, \dots\}$ proporciona una representación matemática del tiempo meteorológico en Ferrol a lo largo de los días.

Capítulo 3

Definiciones básicas de las cadenas de Markov

Antes de comenzar, vamos a ponernos en contexto con un ejemplo.

Ejemplo 3.1. Consideremos que en un locutorio telefónico con 5 líneas de teléfono, en un instante de tiempo dado puede haber un número cualquiera de líneas ocupadas. Durante un período de tiempo se observan las líneas telefónicas a intervalos de 2 minutos y se anota el número de líneas ocupadas en cada instante.

- Sea X_1 la variable aleatoria que representa el número de líneas ocupadas en el instante $t = 1$ (minuto 1), sea X_2 la variable aleatoria que representa el número de líneas ocupadas en el instante $t = 2$ minutos y en general, para $n = 1, 2, \dots$, X_n es una variable aleatoria que representa el número de líneas ocupadas en el instante n -ésimo.
- El estado del proceso en cualquier instante de tiempo es el número de líneas que están siendo utilizadas en ese instante.
- Un proceso estocástico como el que acabamos de describir es un proceso en tiempo discreto, ya que las líneas se observan en puntos discretos a lo largo del tiempo.

Para tener una cadena de Markov es necesario que la probabilidad de cada posible número de líneas ocupadas en el instante $n + 1$, dependa solamente del número de líneas ocupadas en el instante n . Así, podemos construir un modelo suponiendo que el número de líneas ocupadas en el instante $n + 1$, X_{n+1} , conocidos los valores de X_0, X_1, \dots, X_n , solo depende del número de líneas ocupadas en el instante n , X_n , y no depende de cómo estuviesen de ocupadas las líneas en los instantes anteriores al instante presente.

Tal supuesto define una cadena de Markov, que es un tipo de proceso estocástico que definiremos a continuación formalmente:

Definición 3.2. Un proceso estocástico en tiempo discreto $\{X_n, n = 0, 1, \dots\}$ con espacio de estados discreto, E , es una **cadena de Markov** si

$$\mathbb{P}\{X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = i\} = \mathbb{P}\{X_{n+1} = j | X_n = i\} = p_{ij}$$

para $n = 0, 1, 2, \dots$ y para todo $\{i, j, i_0, i_1, \dots, i_{n-1}\} \subset E$. Esta condición es conocida como la *propiedad de Markov*.

El valor p_{ij} representa la probabilidad de que el proceso, cuando esté en el estado i , “haga una transición” al estado j .

Además, dado un proceso estocástico, si $X_n = i$, decimos que el proceso está en el estado i en el momento n .

Todos los estudios que realizaremos a continuación serán para cadenas de Markov en tiempo discreto, pero las nociones de cadenas de Markov se puede extender a un tiempo continuo aunque no vamos a estudiar este tipo de procesos.

Definición 3.3. Una cadena de Markov con espacio de estados $E = \{0, 1, 2, \dots\}$, se dice **homogénea** si, para todo $m, n \in \{0, 1, \dots\}$ y para todo $i, j \in E$ se tiene que:

$$\mathbb{P}\{X_{m+n} = j | X_n = i\} = \mathbb{P}\{X_m = j | X_0 = i\}$$

A lo largo del trabajo consideraremos sólo cadenas de Markov con probabilidades de transición “homogéneas en el tiempo”, es decir, asumimos que $\mathbb{P}\{X_{m+1} = j | X_m = i\} = p_{ij}, \forall i, j \in E$ independientemente del parámetro m .

Definición 3.4. Las probabilidades condicionadas $\mathbb{P}\{X_{n+1} = j | X_n = i\} = \mathbb{P}\{X_n = j | X_{n-1} = i\} = p_{ij}$ de una cadena de Markov se llaman **probabilidades de transición** (de un solo paso).

Definición 3.5. Cuando estas probabilidades de transición son independientes del tiempo $n, \forall n$, decimos que la cadena de Markov tiene **probabilidades de transición estacionarias**.

En esta situación, toda cadena de Markov $\{X_n, n = 0, 1, \dots\}$ está completamente determinada por la distribución de probabilidad del estado inicial $X_0, \mathbb{P}[X_0 = i], i = 0, 1, 2, \dots$, y las probabilidades de transición de un solo paso p_{ij} .

Definición 3.6. Llamamos **probabilidades de transición en n pasos** a:

$$\mathbb{P}\{X_{m+n} = j | X_m = i\} = \mathbb{P}\{X_n = j | X_0 = i\} \text{ para todo } n = 0, 1, 2, \dots$$

donde $i, j \in E = \{0, 1, 2, \dots\}$ y $m = 0, 1, 2, \dots$.

Las probabilidades de transición en n pasos son simplemente las probabilidades condicionadas de que el sistema se encuentre en el estado j exactamente después de n pasos, dado que comenzó en el estado i en un instante m .

En este caso usaremos la notación $p_{ij}^{(n)} = \mathbb{P}\{X_{m+n} = j | X_m = i\}$. Como se puede observar, se tiene que $p_{ij}^{(1)} = p_{ij}$.

Dado que las $p_{ij}^{(n)}$ son probabilidades y que el proceso debe hacer una transición a algún estado, tenemos entonces que, para $n = 0, 1, \dots$:

$$p_{ij}^{(n)} \geq 0 \quad i, j \geq 0; \quad \sum_{j=0}^{\infty} p_{ij}^{(n)} = 1 \quad i = 0, 1, \dots$$

Definición 3.7. Llamamos **matriz de transición en n pasos** a la matriz donde cada elemento (i, j) es la probabilidad de transición $p_{ij}^{(n)}$ en n pasos del estado i al estado j .

$$\mathbf{P}^{(n)} = \begin{array}{c} \begin{array}{cccc} & 0 & 1 & 2 & \dots \end{array} \\ \begin{array}{l} 0 \\ 1 \\ \vdots \\ i \\ \vdots \end{array} \end{array} \begin{bmatrix} p_{00}^{(n)} & p_{01}^{(n)} & p_{02}^{(n)} & \dots \\ p_{10}^{(n)} & p_{11}^{(n)} & p_{12}^{(n)} & \dots \\ \vdots & \vdots & \vdots & \vdots \\ p_{i0}^{(n)} & p_{i1}^{(n)} & p_{i2}^{(n)} & \dots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Cuando $n = 1$ llamaremos \mathbf{P} a la matriz de transición.

Ejemplo 3.8. Volviendo al ejemplo del clima que se presentó en la sección anterior, recordemos que la evolución del tiempo día tras día en Ferrol se ha formulado como un proceso estocástico $\{X_n, n = 0, 1, 2, \dots\}$ donde:

$$X_n = \begin{cases} 0 & \text{si no llueve} \\ 1 & \text{si llueve} \end{cases}$$

Con probabilidades dadas por:

$$\mathbb{P}\{X_{n+1} = 0 | X_n = 0\} = 0'8,$$

$$\mathbb{P}\{X_{n+1} = 0 | X_n = 1\} = 0'6.$$

Además, como ya vimos, estas probabilidades no cambian si también se tiene en cuenta el tiempo antes del día de hoy (día n). En base a esto, obtenemos las siguientes igualdades:

$$\mathbb{P}\{X_{n+1} = 0 | X_0 = j_0, X_1 = j_1, \dots, X_{n-1} = j_{n-1}, X_n = 0\} = \mathbb{P}\{X_{n+1} = 0 | X_n = 0\},$$

$$\mathbb{P}\{X_{n+1} = 0 | X_0 = j_0, X_1 = j_1, \dots, X_{n-1} = j_{n-1}, X_n = 1\} = \mathbb{P}\{X_{n+1} = 0 | X_n = 1\}.$$

para $n = 0, 1, \dots$ y $\{j_0, j_1, \dots, j_{n-1}\} \subset E$.

Estas ecuaciones también deben cumplirse si cambiamos $X_{n+1} = 0$ por $X_{n+1} = 1$ (la razón es que los estados 0 y 1 son mutuamente excluyentes, son los únicos estados posibles y además, las probabilidades de los dos estados deben sumar 1). Por lo tanto, el proceso estocástico tiene la propiedad de Markov, lo que lo convierte en una cadena de Markov. Si usamos la notación que se introdujo en esta sección, las probabilidades de transición (de un paso) son

$$p_{00} = \mathbb{P}\{X_{n+1} = 0 | X_n = 0\} = 0'8,$$

$$p_{10} = \mathbb{P}\{X_{n+1} = 0 | X_n = 1\} = 0'6.$$

para toda $n = 1, 2, \dots$, por lo que éstas son las probabilidades de transición estacionarias. Además:

$$p_{00} + p_{01} = 1 \implies p_{01} = 1 - 0'8 = 0'2,$$

$$p_{10} + p_{11} = 1 \implies p_{11} = 1 - 0'6 = 0'4.$$

Por lo tanto, la matriz de transición viene dada por

$$\mathbf{P} = \begin{array}{c} 0 \\ 1 \end{array} \begin{array}{cc} 0 & 1 \\ \left[\begin{array}{cc} p_{00} & p_{01} \\ p_{10} & p_{11} \end{array} \right] \end{array} = \begin{array}{c} 0 \\ 1 \end{array} \begin{array}{cc} 0 & 1 \\ \left[\begin{array}{cc} 0'8 & 0'2 \\ 0'6 & 0'4 \end{array} \right] \end{array}$$

donde estas probabilidades se refieren a la transición del estado i de la fila al estado j de la columna. Recordemos que el estado es 0 si no llueve y 1 si llueve, así que estas probabilidades de transición proporcionan la probabilidad del estado del tiempo el día $n + 1$ dado el estado del tiempo el día n .

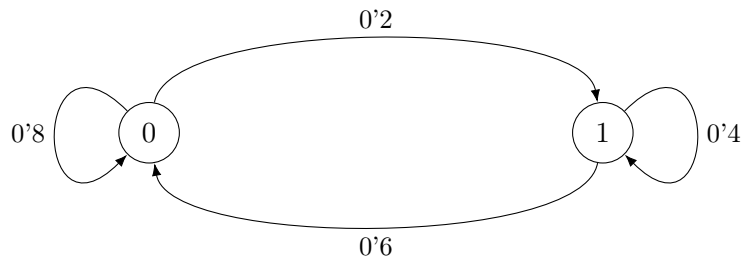


Figura 3.1: Grafo que modeliza el ejemplo.

El diagrama de transición de estado de la figura 3.1 presenta de manera gráfica la misma información que proporciona la matriz de transición. Los vértices representan los dos estados posibles del clima, mientras que las aristas muestran las transiciones posibles de un día al siguiente (se incluye la opción de regresar al mismo estado). Cada una de las probabilidades de transición se indican en la arista correspondiente.

Capítulo 4

Ecuaciones de Chapman-Kolmogorov

La ecuación de Chapman-Kolmogorov fue derivada de forma independiente tanto por el matemático británico Sydney Chapman como por el matemático ruso Andrey Kolmogorov. En matemáticas, específicamente en teoría de probabilidad, y en particular, en la teoría de procesos estocásticos markovianos estas ecuaciones son muy útiles.

En la sección anterior ya introducimos la probabilidad de transición de n pasos $p_{ij}^{(n)}$, las ecuaciones de Chapman-Kolmogórov lo que nos proporcionan es un método para calcular estas probabilidades de transición de n pasos.

Teorema 4.1 (Ecuaciones de Chapman-Kolmogórov). *Dado el espacio de estados $E = \{0, 1, \dots\}$, se cumple que para todo $n, m = 0, 1, \dots$*

$$p_{ij}^{(n+m)} = \sum_{k \in E} p_{ik}^{(n)} p_{kj}^{(m)} \quad (4.1)$$

con $i, j \in E$.

Demostración. Una prueba formal es la siguiente. Al condicionar el estado de Markov cadena en el tiempo $t = n$, encontramos

$$\begin{aligned} \underbrace{\mathbb{P}\{X_{n+m} = j | X_0 = i\}}_{p_{ij}^{(n+m)}} &= \sum_{k \in E} \mathbb{P}\{X_{n+m} = j | X_0 = i, X_n = k\} \mathbb{P}\{X_n = k | X_0 = i\} \\ &= \sum_{k \in E} \mathbb{P}\{X_{n+m} = j | X_n = k\} \mathbb{P}\{X_n = k | X_0 = i\} \\ &= \sum_{k \in E} \underbrace{\mathbb{P}\{X_m = j | X_0 = k\}}_{p_{kj}^{(m)}} \underbrace{\mathbb{P}\{X_n = k | X_0 = i\}}_{p_{ik}^{(n)}}. \end{aligned}$$

que verifica la ecuación 4.1. Tenemos que tener en cuenta que la segunda igualdad usa la propiedad de Markov y la última igualdad utiliza el supuesto de que la cadena de Markov es homogénea en el tiempo. \square

El teorema establece que la probabilidad de ir de i a j en $(n + m)$ pasos se obtiene sumando las probabilidades de los eventos mutuamente excluyentes de pasar primero del estado i a algún estado k en n pasos y luego pasar del estado k al estado j en m pasos.

NOTA: En particular, tenemos para cualquier $n = 1, 2, \dots$

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

Por tanto, las probabilidades de transición de n pasos $p_{ij}^{(n)}$ se pueden calcular recursivamente a partir de las probabilidades de transición de un paso p_{ij} .

Si denotamos por $\mathbf{P}^{(n)}$ a la matriz de probabilidades de transición de n pasos, entonces la ecuación 4.1 afirma que $\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \cdot \mathbf{P}^{(m)}$.

En particular, $\mathbf{P}^{(2)} = \mathbf{P}^{(1+1)} = \mathbf{P} \cdot \mathbf{P} = \mathbf{P}^2$ y por inducción $\mathbf{P}^{(n)} = \mathbf{P}^{(n-1+1)} = \mathbf{P}^{n-1} \cdot \mathbf{P} = \mathbf{P}^n$.

Entonces, los $p_{ij}^{(n)}$ son los elementos de la matriz \mathbf{P}^n . Y \mathbf{P} denota la matriz cuyo elemento (i, j) es la probabilidad de transición de un paso p_{ij} .

Ejemplo 4.2. Volviendo al ejemplo del tiempo en Ferrol, vamos a usar las fórmulas que acabamos de dar para obtener las matrices de transición de n pasos a partir de la matriz de transición \mathbf{P} de un paso que ya hemos calculado:

$$\mathbf{P}^{(2)} = \mathbf{P} \cdot \mathbf{P} = \begin{bmatrix} 0'8 & 0'2 \\ 0'6 & 0'4 \end{bmatrix} \begin{bmatrix} 0'8 & 0'2 \\ 0'6 & 0'4 \end{bmatrix} = \begin{bmatrix} 0'76 & 0'24 \\ 0'72 & 0'28 \end{bmatrix}$$

Y esto lo interpretamos como que si el tiempo está en el estado 0 un día concreto, la probabilidad de que de que el tiempo esté en el estado 0 dos días después es 0'76 y de que esté en el estado 1 dos días después es de 0'24.

Análogamente si el tiempo está en el estado 1, la probabilidad de que esté en el estado 0 dos días después es 0'72 y de que esté en el estado 1 dos días después es 0'28.

Las probabilidades del estado del clima a 3, 4, 5, ... días se pueden calcular de la misma

forma a partir de las matrices de transición de 3, 4, 5, ... pasos:

$$\mathbf{P}^{(3)} = \mathbf{P}^3 = \mathbf{P} \cdot \mathbf{P}^2 = \begin{bmatrix} 0'8 & 0'2 \\ 0'6 & 0'4 \end{bmatrix} \begin{bmatrix} 0'76 & 0'24 \\ 0'72 & 0'28 \end{bmatrix} = \begin{bmatrix} 0'752 & 0'248 \\ 0'744 & 0'256 \end{bmatrix}$$

$$\mathbf{P}^{(4)} = \mathbf{P}^4 = \mathbf{P} \cdot \mathbf{P}^3 = \begin{bmatrix} 0'8 & 0'2 \\ 0'6 & 0'4 \end{bmatrix} \begin{bmatrix} 0'752 & 0'248 \\ 0'744 & 0'256 \end{bmatrix} = \begin{bmatrix} 0'75 & 0'25 \\ 0'749 & 0'251 \end{bmatrix}$$

$$\mathbf{P}^{(5)} = \mathbf{P}^5 = \mathbf{P} \cdot \mathbf{P}^4 = \begin{bmatrix} 0'8 & 0'2 \\ 0'6 & 0'4 \end{bmatrix} \begin{bmatrix} 0'75 & 0'25 \\ 0'749 & 0'251 \end{bmatrix} = \begin{bmatrix} 0'75 & 0'25 \\ 0'75 & 0'25 \end{bmatrix}$$

$$\mathbf{P}^{(6)} = \mathbf{P}^6 = \mathbf{P} \cdot \mathbf{P}^5 = \begin{bmatrix} 0'8 & 0'2 \\ 0'6 & 0'4 \end{bmatrix} \begin{bmatrix} 0'75 & 0'25 \\ 0'75 & 0'25 \end{bmatrix} = \begin{bmatrix} 0'75 & 0'25 \\ 0'75 & 0'25 \end{bmatrix}$$

⋮

Como podemos observar a partir de $\mathbf{P}^{(5)}$ la probabilidad “se estabiliza” y esto muestra que la probabilidad de que el tiempo esté en un estado no va a depender del tiempo 5 días antes.

A esto lo vamos a denominar probabilidades del estado estable, las cuales definiremos y estudiaremos más adelante.

Capítulo 5

Tiempos de primera pasada

Muchas aplicaciones de las cadenas de Markov involucran cadenas en las que algunos de los estados son “absorbentes” y los otros estados son “transitorios”.

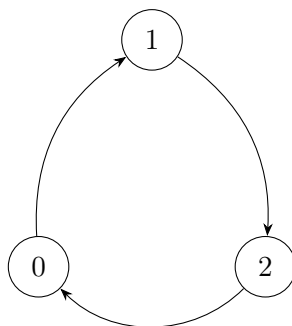
Para definir estos conceptos de estados transitorios y estados recurrentes, introduciremos los tiempos de primera pasada con sus probabilidades correspondientes.

Definición 5.1. Llamamos **tiempos de primera pasada**, al número de transiciones que hace el proceso al ir del estado i al estado j por primera vez.

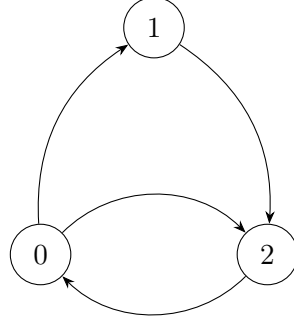
Definición 5.2. En el caso $i = j$, el tiempo de primera pasada es el número de transiciones hasta que el proceso regresa al estado inicial i . En este caso, el tiempo de primera pasada se llama **tiempo de recurrencia** del estado i .

Ejemplo 5.3. A continuación vamos a ilustrar un caso básico de tiempos de primera pasada donde no hay distribuciones de probabilidad que intervengan en el cálculo, y solo dependen de los pasos necesarios para llegar de un estado al otro.

¿Cuántos pasos se necesitan para pasar de 0 a 2? En este caso podemos ver que se necesitan 2 pasos. Así el tiempo de primera pasada de ir del estado 0 al 2 es 2.



Otro ejemplo sería el siguiente, donde se necesitan 1 ó 2 pasos para ir del estado 0 al 2 y por tanto el tiempo de primera pasada de ir del estado 0 al 2 será 1 ó 2.



Definición 5.4. Dada una cadena de Markov en tiempo discreto con espacio de estados $E = \{0, 1, 2, \dots\}$ y probabilidades de transición de un paso p_{ij} , donde $i, j \in E$. Para $n = 1, 2, \dots$, se define **probabilidad del tiempo de primera pasada**, $f_{ij}^{(n)}$, la definimos como:

$$f_{ij}^{(n)} = \mathbb{P}\{X_n = j, X_k \neq k \text{ para } 1 \leq k \leq n-1 | X_0 = i\} \quad i, j \in E$$

En otras palabras, $f_{ij}^{(n)}$ denota la probabilidad de que la primera transición del proceso al estado j sea en el tiempo n cuando el proceso empieza en el estado i .

Llamaremos:

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

Observemos que $f_{ij} = \mathbb{P}\{X_n = j \text{ para algún } n \geq 1 | X_0 = i\}$, es decir, f_{ij} denota la probabilidad de que el proceso, en algún momento, haga la transición al estado j cuando el proceso empezó en el estado i .

Proposición 5.5. *Se pueden obtener las siguientes relaciones recursivas:*

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

⋮

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

Entonces, la probabilidad de un tiempo de primera pasada del estado i al estado j en n pasos, se puede calcular de manera recursiva a partir de las probabilidades de transición de un paso.

Demostración. Para empezar tenemos que

$$f_{ij}^{(1)} = p_{ij}^{(1)} = p_{ij} \text{ ya que en un paso siempre es por primera vez.}$$

Además $p_{ij}^{(2)} = f_{ij}^{(2)} + f_{ij}^{(1)} p_{jj}$ por el teorema de las probabilidades totales ya que hay dos formas de pasar de i a j en dos pasos:

- Primera forma: Estar en el estado i en X_0 , estar en un estado distinto de j en X_1 y estar en el estado j en X_2 , cuya probabilidad es $f_{ij}^{(2)}$.
- Segunda forma: Estar en el estado i en X_0 , y estar en el estado j en X_1 y X_2 , cuya probabilidad es $f_{ij}^{(1)} p_{jj}$

Como estas probabilidades son mutuamente excluyentes y exhaustivas, podemos aplicar el teorema de la probabilidad total obteniendo que $p_{ij}^{(2)} = f_{ij}^{(2)} + f_{ij}^{(1)} p_{jj}$ como acabamos de decir.

De donde podemos despejar $f_{ij}^{(2)} = p_{ij}^{(2)} - f_{ij}^{(1)} p_{jj}$.

Análogamente podemos hacer esto para n pasos, aplicando el teorema de probabilidad total considerando los siguientes sucesos mutuamente excluyentes y exhaustivos:

“Que la primera transición del proceso al estado j sea en el tiempo $1, 2, \dots, n$, cuando el proceso empieza en el estado i ”. Así:

$$p_{ij}^{(n)} = f_{ij}^{(n)} + f_{ij}^{(1)} p_{jj}^{(n-1)} + f_{ij}^{(2)} p_{jj}^{(n-2)} + \dots + f_{ij}^{(n-1)} p_{jj}$$

Y de aquí podemos despejar

$$f_{ij}^{(n)} = p_{ij}^{(n)} - f_{ij}^{(1)} p_{jj}^{(n-1)} - f_{ij}^{(2)} p_{jj}^{(n-2)} - \dots - f_{ij}^{(n-1)} p_{jj} = p_{ij}^{(n)} - \sum_{r=1}^{n-1} f_{ij}^{(r)} p_{jj}^{(n-r)}$$

□

NOTA: Dados i, j fijos, las $f_{ij}^{(n)}$ son números reales no negativos tales que

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

Si la suma es estrictamente menor que uno entonces el proceso que se inicia en un estado i puede que no alcance el estado j .

Si la suma es igual a uno, lo que tenemos es $\{f_{ij}^{(1)}, f_{ij}^{(2)}, \dots, f_{ij}^{(n)}, \dots\}$, la distribución de la variable aleatoria “tiempo de primera pasada de i a j ”.

Ejemplo 5.6. Supongamos un modelo de predicción del tiempo, modelizado por una cadena de Markov con los siguientes estados:

Estado 0: Día soleado.

Estado 1 : Día lluvioso

Estado 2 : Día nublado.

El clima del día siguiente depende únicamente del clima del día actual y no del clima de los días anteriores. Si el día actual es soleado, el día siguiente será soleado, nublado o lluvioso con probabilidades respectivas de 0'3, 0'2 y 0'5. Las probabilidades de transición son 0'6, 0'4 y 0 cuando el día actual está nublado y son 0'2, 0'7 y 0'1 cuando el día actual es lluvioso. Así, la matriz \mathbf{P} de probabilidades de transición de un paso p_{ij} es dada por:

$$\mathbf{P} = \begin{bmatrix} 0'3 & 0'2 & 0'5 \\ 0'6 & 0'4 & 0 \\ 0'2 & 0'7 & 0'1 \end{bmatrix}$$

Usando la proposición anterior, vamos a ver como obtener los valores de la función de probabilidad de los tiempos de primera pasada del estado 1 al 2 , para $n=1,2,3,4$.

Necesitamos las siguientes matrices de transición:

$$\mathbf{P}^2 = \begin{bmatrix} 0'31 & 0'49 & 0'2 \\ 0'42 & 0'28 & 0'3 \\ 0'5 & 0'39 & 0'11 \end{bmatrix}, \mathbf{P}^3 = \begin{bmatrix} 0'427 & 0'398 & 0'175 \\ 0'354 & 0'406 & 0'24 \\ 0'406 & 0'333 & 0'261 \end{bmatrix}, \mathbf{P}^4 = \begin{bmatrix} 0'4019 & 0'3671 & 0'231 \\ 0'3978 & 0'4012 & 0'201 \\ 0'3738 & 0'3971 & 0'2291 \end{bmatrix}$$

Tenemos que:

$$f_{12}^{(1)} = p_{12} = 0$$

$$f_{12}^{(2)} = p_{12}^{(2)} - p_{22}f_{12}^{(1)} = 0'3 - 0(0'1) = 0'3$$

$$f_{12}^{(3)} = p_{12}^{(3)} - p_{22}^{(2)}f_{12}^{(1)} - p_{22}f_{12}^{(2)} = 0'24 - 0(0'11) - 0'3(0'1) = 0'21$$

$$f_{12}^{(4)} = p_{12}^{(4)} - p_{22}^{(3)}f_{12}^{(1)} - p_{22}^{(2)}f_{12}^{(2)} - p_{22}f_{12}^{(3)} = 0'201 - 0(0'261) - 0'3(0'11) - 0'21(0'1) = 0'147$$

donde p_{1k} y $f_{12}^{(1)}$ se obtienen de la matriz de transición de un paso.

Los $p_{1k}^{(2)}, p_{1k}^{(3)}, p_{1k}^{(4)}$ se obtienen de las matrices de transición de 2, 3 y 4 pasos.

Finalizamos este capítulo dando dos definiciones que están relacionadas con algunos conceptos que veremos más adelante.

Definición 5.7. Llamamos **tiempo de absorción de un estado j** a $\tau = \min\{n \geq 1 | X_n = j\}$

Definición 5.8. Llamamos **tiempo esperado de primera pasada** a μ_{ij} y lo definimos como:

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

Capítulo 6

Clasificación de los estados

Consideremos una cadena de Markov homogénea con espacio de estados $E = \{0, 1, \dots\}$ y matriz de transición \mathbf{P} . Ya sabemos que p_{ij} es la probabilidad de un solo paso de que haya una transición de i a j .

Antes de hablar de los tipos de estados en una cadena de Markov, vamos a introducir algunas definiciones necesarias.

Definición 6.1. Se dice que el estado j es accesible desde el estado i si $p_{ij}^{(n)} > 0$ para alguna $n \geq 0$.

Entonces, que el estado j sea accesible desde el estado i quiere decir que el proceso puede llegar al estado j empezando en el estado i .

Ejemplo 6.2. En el ejemplo del clima del que hablamos en las secciones anteriores es fácil ver que dos estados cualquiera i, j son accesibles ya que $p_{ij} > 0$ para toda i, j .

Definición 6.3. Dados dos estados i y j , decimos que se **comunican** si el estado j es accesible desde el estado i y el estado i es accesible desde el estado j .

Propiedades

Se tiene que:

- Cualquier estado se comunica consigo mismo (porque $p_{ii}^{(0)} = \mathbb{P}\{X_0 = i | X_0 = i\} = 1$).
- Si el estado i se comunica con el estado j , entonces el estado j se comunica con el estado i .
- Si el estado i se comunica con el estado j y este con el estado k , entonces el estado i se comunica con el estado k .

Las propiedades 1 y 2 se deducen de la definición de estados que se comunican, mientras que la propiedad 3 se deriva de las ecuaciones de Chapman-Kolmogorov.

Definición 6.4. Se dice que la cadena de Markov es **irreducible** si todos los estados se comunican.

Ejemplo 6.5. Dada la cadena de Markov con estados 0, 1 y 2 y con matriz de transición dada por:

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

Es fácil verificar que la cadena de Markov es irreducible. Por ejemplo, es posible pasar del estado 0 al estado 2 ya que podemos empezar en el estado 0, ir al estado 1 (con probabilidad $\frac{1}{2}$) y una vez estamos en el estado 1 pasamos al estado 2 (con probabilidad $\frac{1}{4}$).

Algo útil en el estudio de las cadenas de Markov es saber si dado un proceso que comienza en un estado, regresará alguna vez a él. Basándonos en esto, vamos a empezar la clasificación hablando de estados transitorios y recurrentes.

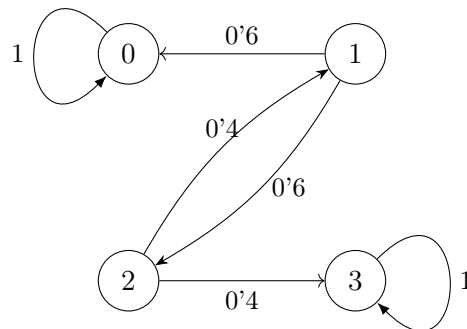
Estados transitorios y recurrentes

Definición 6.6. Un estado se llama **estado transitorio** si $f_{ii} < 1$.

Ejemplo 6.7. Supongamos una cadena de Markov con espacio de estados discreto y finito $E = \{0, 1, 2, 3\}$ y con probabilidades de transición dadas por la siguiente matriz de transición:

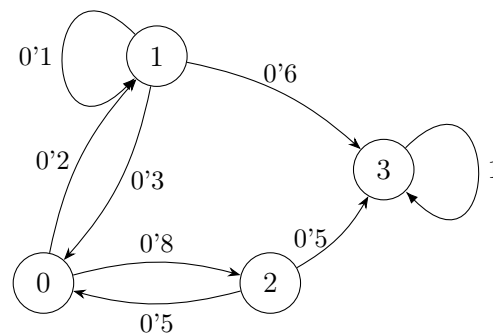
$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0'6 & 0 & 0'4 & 0 \\ 0 & 0'6 & 0 & 0'4 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Podemos ver que los estado 1 y 2 son transitorios, ya que el proceso los acabará abandonando para entrar en el estado 0 o en el 3 y después permanecerá ahí de manera indefinida. Esto lo podemos ver claramente en el grafo de esta cadena de Markov:



Definición 6.8. Llamamos **estado recurrente** a aquel que cumple que $f_{ii} = 1$, es decir, si el proceso comienza en un estado i , en algún momento la cadena volverá al estado i .

Ejemplo 6.9. Dada la siguiente cadena de Markov :



Con matriz de transición:

$$\mathbf{P} = \begin{bmatrix} 0 & 0'2 & 0'8 & 0 \\ 0'3 & 0'1 & 0 & 0'6 \\ 0'5 & 0 & 0 & 0'5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Es claro que el estado 3 es recurrente ya que $p_{33} = 1$, por tanto tomando el estado inicial 3, $X_0 = 3$, con probabilidad 1 se vuelve al estado 3.

A continuación daremos unos resultados útiles para saber si un estado es transitorio o recurrente.

Propiedades

- Para cualquier estado i , denotamos por f_{ii} la probabilidad de que, comenzando en el estado i , el proceso regrese a él.

Como acabamos de ver si $f_{ii} < 1$ el estado i será transitorio y si $f_{ii} = 1$ el estado i será recurrente.

Suponemos que el proceso comienza en un estado i que es recurrente. Por tanto, con probabilidad 1, el proceso en algún momento volverá al estado i . Sin embargo, según la definición de cadena de Markov, se deduce que el proceso comenzará de nuevo cuando vuelva a entrar en el estado i , y por lo tanto, el estado i será visitado nuevamente en algún momento.

Repitiendo continuamente este argumento llegamos a que si el estado i es recurrente, entonces comenzando en este estado i , el proceso volverá a entrar en el estado i una y otra vez.

Por otro lado si suponemos que el estado i es transitorio, cada vez que el proceso entra en el estado i habrá una probabilidad positiva, $1 - f_{ii}$, de que no vuelva a entrar a ese estado.

Por lo tanto, comenzando en el estado i , la probabilidad de que el proceso vuelva al estado i al cabo de exactamente n pasos es igual a $f_{ii}^{n-1}(1 - f_{ii})$, $n \geq 1$.

En otras palabras, si i es transitorio, entonces, comenzando en el estado i , el número de períodos de tiempo hasta que el proceso vuelve al estado i tiene una distribución geométrica con media finita $\frac{1}{1-f_{ii}}$.

A partir de esta ideas ideas, se sigue que el estado i es recurrente si y solo si, comenzando en el estado i , el número esperado de períodos de tiempo en que el proceso está en el estado i es infinito. Pero tomando

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

tenemos que $\sum_{n=0}^{\infty} I_n$ representa el número de períodos en los que el proceso está en el estado i .

Además

$$\mathbb{E}\left(\sum_{n=0}^{\infty} I_n | X_0 = i\right) = \sum_{n=0}^{\infty} \mathbb{E}(I_n | X_0 = i) = \sum_{n=0}^{\infty} \mathbb{P}\{X_n = i | X_0 = i\} = \sum_{n=0}^{\infty} p_{ii}^{(n)}.$$

Con todo esto podemos afirmar que:

Un estado es transitorio si y solo si $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.

Un estado i es recurrente si y solo si $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.

- Además estos argumentos también nos muestran que un estado transitorio solo será visitado un número finito de veces. Esto lleva a la conclusión de que en una cadena de Markov de estados finitos no todos los estados pueden ser transitorios.

Para ver esto, suponemos que los estados $0, 1, \dots, N$ son transitorios. Así, después de una cantidad de tiempo finita (sea después de un tiempo t_0), el estado 0 nunca será visitado, y después de un tiempo, t_1 , el estado 1 nunca será visitado, etc.

Entonces, después de un tiempo finito $t = \max\{t_0, t_1, \dots, t_N\}$ no se visitará ningún estado.

Pero como el proceso debe estar en un estado después de un tiempo t , llegamos a una contradicción, lo que nos muestra que al menos un estado debe ser recurrente.

Lema 6.10. *Suponemos que j es transitorio. Entonces para cualquier estado $i \in E$ se tiene que*

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

Demostración. Tomando ahora un estado inicial i tal que $i \neq j$. Dada la siguiente relación obtenida en la proposición 5.5 :

$$p_{ij}^{(n)} = \sum_{k=1}^n p_{jj}^{(n-k)} f_{ij}^{(k)}, n = 1, 2, \dots$$

Ahora, sumando n , y usando que $p_{jj}^{(1)} = 1$, obtenemos:

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

Como ya vimos si j transitorio, entonces $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$ y como $f_{ij} \leq 1$ y se obtiene $\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty$ y así $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = 0$. \square

Corolario 6.11. Si i y j son dos estados que se comunican entre sí y j es recurrente, entonces i también es recurrente.

Demostración. Para probar esto primero tenemos que tener en cuenta que como el estado i se comunica con el estado j , entonces existen enteros k y m tales que $p_{ij}^k > 0$ y $p_{ji}^m > 0$. Ahora para cualquier entero n se tiene que:

$$p_{jj}^{m+n+k} \geq p_{ji}^m p_{ii}^n p_{ij}^k$$

Esto se sigue ya que p_{jj}^{m+n+k} es la probabilidad de ir del estado j al estado j en $m+n+k$ pasos, mientras que $p_{ji}^m p_{ii}^n p_{ij}^k$ es la probabilidad de ir del estado j al estado j en $m+n+k$ pasos a través de una ruta que va de j a i en m pasos, luego de i a i en n pasos adicionales, y por último de i a j en k pasos adicionales.

De lo anterior obtenemos, sumando sobre n , que

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

ya que $p_{ji}^m p_{ij}^k > 0$ y $\sum_{n=1}^{\infty} p_{ii}^n$ es infinito si i es recurrente. Por tanto, por la proposición anterior se sigue que j también es recurrente. \square

Estados absorbentes

En esta sección vamos a hablar de un caso especial de un estado recurrente que recibe el nombre de estado absorbente y que definimos a continuación:

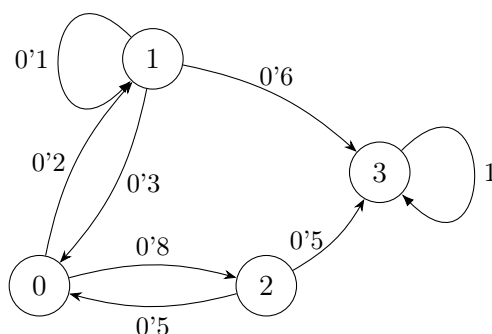
Definición 6.12. Un estado se llama **estado absorbente** si, después de haber entrado ahí, el proceso nunca saldrá de él. En consecuencia, el estado i es un estado absorbente si y sólo si $p_{ii} = 1$.

Como ya dijimos, todos los estados absorbentes son recurrentes ya que si i es un estado absorbente $p_{ii}^{(1)} = p_{ii} = 1$ y como $f_{ii}^{(1)} = p_{ii}^{(1)}$ entonces $f_{ii}^{(1)} = 1$, mientras que $p_{ii}^{(n)} = 0$ para $n = 2, 3, \dots$

Por otro lado, $f_{ii}^{(n)} = p_{ii}^{(n)} - \sum_{r=1}^{(n-1)} f_{ii}^{(r)} p_{ii}^{(n-r)}$ y usando las probabilidades que acabamos de dar llegamos a que $f_{ii}^{(n)} = 0$ para $n = 2, 3, \dots$

Por último usando que $f_{ii} = \sum_{k=1}^{\infty} f_{ii}^{(k)}$ y todo lo anterior llegamos a la conclusión de que $f_{ii} = 1$ por lo que i es recurrente.

Ejemplo 6.13. Como ya vimos en el ejemplo 6.9, en la cadena de Markov dada por:



tenemos que el estado 3 era recurrente. Y podemos ver que este mismo estado es absorbente (caso particular de un estado recurrente) ya que una vez que se llega a este estado no se sale de él.

Hay veces que existen dos o más estados absorbentes en una cadena de Markov, entonces el proceso será absorbido en alguno de esos estados. Por tanto nos será de utilidad calcular las probabilidades de absorción que definiremos a continuación:

Definición 6.14. Si k es un estado absorbente y el proceso comienza en el estado i , la probabilidad de llegar en algún momento a k se llama **probabilidad de absorción** al estado k , dado que el sistema comenzó en el estado i . Observemos que esta probabilidad es f_{ik} .

Estados periódicos y aperiódicos

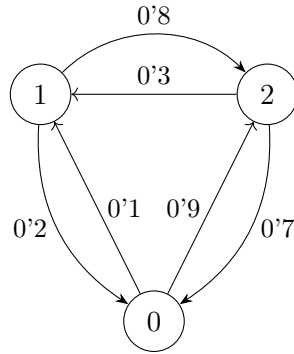
Otra propiedad útil de las cadenas de Markov es la periodicidad, y por eso vamos a introducir las siguientes definiciones y hacer una clasificación de los estados en base a estas.

Definición 6.15. El estado i tiene **período** d si $p_{ii}^{(n)} = 0$ cuando n no es divisible por d y d es el mayor entero con esa propiedad, es decir, d es el máximo común divisor del conjunto $\{n \geq 0 : p_{ii}^{(n)} > 0\}$.

Definición 6.16. Decimos que un estado es **periódico** si, partiendo de ese estado, sólo es posible volver a él en un número de etapas que sea múltiplo de un cierto número entero mayor que uno.

Mientras que decimos que un estado es **aperiódico** si tiene período 1.

Ejemplo 6.17. Dada la siguiente cadena de Markov, vamos a calcular su periodo.



con matriz de transición dada por:

$$\mathbf{P} = \begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{array}{c} 0 \quad 1 \quad 2 \\ \left[\begin{array}{ccc} 0 & 0'1 & 0'9 \\ 0'2 & 0 & 0'8 \\ 0'3 & 0 & 0'7 \end{array} \right] \end{array}$$

Tenemos el siguientes conjunto:

$$\{n \geq 1 : p_{11}^{(n)} > 0\} = \{2, 3, 4, \dots\},$$

que se obtiene al calcular las siguientes matrices de transición

$$\mathbf{P}^2 = \begin{bmatrix} 0'29 & 0 & 0'71 \\ 0'24 & 0'02 & 0'74 \\ 0'21 & 0'03 & 0'76 \end{bmatrix}, \mathbf{P}^3 = \begin{bmatrix} 0'213 & 0'029 & 0'758 \\ 0'226 & 0'024 & 0'75 \\ 0'234 & 0'021 & 0'745 \end{bmatrix}$$

$$\mathbf{P}^4 = \begin{bmatrix} 0'2332 & 0'0213 & 0'7455 \\ 0'2298 & 0'0226 & 0'7476 \\ 0'2277 & 0'0234 & 0'7489 \end{bmatrix}$$

⋮

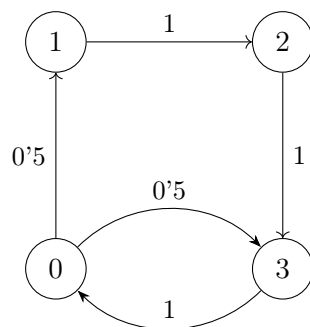
así, vemos que el estado 1 tienen periodo 1 ya que el periodo es el máximo común divisor del conjunto $\{n \geq 1 : p_{11}^{(n)} > 0\}$ y por definición es aperiódico.

Proposición 6.18. *Sea un estado i que tiene periodo t y dado j un estado que se comunica con i , entonces j tiene periodo t .*

Demostración. Asumimos que el estado i tiene un periodo t_i , y que el estado j se comunica con i , sea $n \in \{m \geq 1 : p_{jj}^m > 0\}$. Como i y j se comunican, existen $k, l \neq 1$ tal que $p_{ij}^k > 0$ y $p_{ji}^l > 0$ y usando que $\mathbb{P}(X_n = k | X_0 = i) \geq 0$ se tiene que $p_{ii}^{k+l} > 0$ siendo $k+l$ múltiplo de t_i . Similarmente por $\mathbb{P}(X_n = k | X_0 = i) \geq 0$ también se tiene que p_{ii}^{n+k+l} siendo $n+k+l$ múltiplo de t_i lo que implica que $d_j \geq d_i$.

Intercambiando los roles de i y j obtenemos que $d_i \geq d_j$. □

Ejemplo 6.19. Consideramos la siguiente cadena de Markov:



Vamos a calcular los periodos de cada estado de esta cadena de Markov. Y la matriz de transición de esta cadena está dada por:

$$\mathbf{P} = \begin{matrix} & \begin{matrix} 0 & 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{matrix}$$

Y las matrices de transición de n pasos para $n = 1, 2, 3, 4, 5, 6, \dots$ son

$$\mathbf{P}^2 = \begin{bmatrix} 0'5 & 0 & 0'5 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0'5 & 0 & 0'5 \end{bmatrix}, \mathbf{P}^3 = \begin{bmatrix} 0 & 0'25 & 0 & 0'75 \\ 1 & 0 & 0 & 0 \\ 0 & 0'5 & 0 & 0'5 \\ 0'5 & 0 & 0'5 & 0 \end{bmatrix}$$

$$\mathbf{P}^4 = \begin{bmatrix} 0'75 & 0 & 0'25 & 0 \\ 0 & 0'5 & 0 & 0'5 \\ 0'5 & 0 & 0'5 & 0 \\ 0 & 0'25 & 0 & 0'75 \end{bmatrix}, \mathbf{P}^5 = \begin{bmatrix} 0 & 0'375 & 0 & 0'625 \\ 0'5 & 0 & 0'5 & 0 \\ 0 & 0'25 & 0 & 0'75 \\ 0'75 & 0 & 0'25 & 0 \end{bmatrix}$$

$$\mathbf{P}^6 = \begin{matrix} & & 0 & 1 & 2 & 3 \\ \begin{matrix} 0 \\ 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} 0'625 & 0 & 0'375 & 0 \\ 0 & 0'25 & 0 & 0'75 \\ 0'75 & 0 & 0'25 & 0 \\ 0 & 0'375 & 0 & 0'625 \end{bmatrix} \end{matrix}$$

⋮

Y como consecuencia tenemos los siguientes conjuntos:

$$\{n \geq 1 : p_{00}^{(n)} > 0\} = \{2, 4, 6, 8, 10, \dots\},$$

$$\{n \geq 1 : p_{11}^{(n)} > 0\} = \{4, 6, 8, 10, \dots\},$$

$$\{n \geq 1 : p_{22}^{(n)} > 0\} = \{4, 6, 8, 10, \dots\},$$

$$\{n \geq 1 : p_{33}^{(n)} > 0\} = \{2, 4, 6, 8, 10, \dots\},$$

así, vemos que todos los estados tienen periodo 2 (por ser el periodo el máximo común divisor de estos conjuntos).

Además, que los estados 0 y 3 tengan el mismo periodo se debe a la proposición que acabamos de ver ya que se comunican.

Por último, vamos a dar una definición que caracteriza un tipo de cadenas muy comunes:

Definición 6.20. En una cadena de Markov de estado finito, los estados recurrentes aperiódicos se llaman **ergódicos**.

Se dice que una cadena de Markov es **ergódica** si todos sus estados son ergódicos.

Y con esto finalizamos la clasificación de los estados. Y por último tenemos los estados periódicos y aperiódicos.

Capítulo 7

Propiedades a largo plazo de las cadenas de Markov

Cuando hablamos de una cadena de Markov, nos interesa conocer no solo su comportamiento a “corto plazo” (es decir, después de un número finito de pasos) sino que también nos será de utilidad conocer su comportamiento a “largo plazo”.

Debido a la extensión y complejidad de los comportamientos a “largo plazo”, solo vamos a presentar el resultado más importante para cadenas de Markov irreducibles y ergódicas, que tiene que ver con las probabilidades a largo plazo en el tiempo.

Teorema 7.1. *Dada una cadena de Markov $\{X_n, n = 0, 1, 2, \dots\}$ irreducible y ergódica con probabilidades de transición en un solo paso p_{ij} , entonces existe el $\lim_{n \rightarrow \infty} p_{ij}^{(n)}$ y es independiente del estado inicial i .*

Además, si llamamos $\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)}$ se verifica que

$$\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij}$$

$$\sum_{j=0}^{\infty} \pi_j = 1$$

La demostración la podemos encontrar en [8].

Definición 7.2. Llamamos **probabilidades del estado estable** a los valores

$\pi = (\pi_1, \pi_2, \dots, \pi_j, \dots)$ definidos como $\pi_j = \lim_{n \rightarrow \infty} p_{ij}^{(n)} > 0$.

Las probabilidades π_j también son conocidas como probabilidades a largo plazo o probabilidades estacionarias, y $\pi = (\pi_0, \pi_1, \dots)$ se conoce como la distribución de probabilidad estacionaria de la cadena de Markov.

NOTA: Como $\lim_{n \rightarrow \infty} p_{ij}^{(n)} = \lim_{n \rightarrow \infty} \mathbb{P}\{X_n = j | X_0 = i\} = \pi_j > 0 \quad j = 0, 1, 2, \dots$ esto significa que a “largo plazo” (cuando $n \rightarrow \infty$) la probabilidad de que la cadena de Markov se encuentre en el estado j es aproximadamente π_j sin importar cuál era el estado inicial de la cadena (en $t = 0$).

NOTA: El sistema de ecuaciones anterior también lo podemos expresar en forma matricial de la siguiente manera:

$$\pi = \pi \mathbf{P}$$

donde $\pi = (\pi_0, \pi_1, \dots)$.

Ejemplo 7.3. Volviendo al ejemplo del tiempo en Ferrol que comentamos en secciones anteriores, veamos cuales son sus ecuaciones del estado estable, que podemos obtener usando el teorema anterior.

$$\pi_0 = \pi_0 p_{00} + \pi_1 p_{10},$$

$$\pi_1 = \pi_0 p_{01} + \pi_1 p_{11},$$

$$1 = \pi_0 + \pi_1.$$

y substituyendo las probabilidades, nos quedaría que:

$$\pi_0 = 0'8\pi_0 + 0'6\pi_1 \implies 0'2\pi_0 = 0'6\pi_1,$$

$$\pi_1 = 0'2\pi_0 + 0'4\pi_1 \implies 0'6\pi_1 = 0'2\pi_0,$$

$$1 = \pi_0 + \pi_1.$$

Podemos observar que las dos primeras ecuaciones son redundantes, así que bastaría resolver el sistema formado por la primera y tercera ecuación (o por la segunda y la tercera) y el resultado es el siguiente:

$$\pi_0 = 0'75 \quad \text{y} \quad \pi_1 = 0'25$$

que son las probabilidades del estado estable.

Si comparamos este resultado con los valores de la matriz $\mathbf{P}^{(5)}$ observamos que son los mismos.

Consideremos de nuevo una cadena de Markov $\{X_n, n = 0, 1, \dots\}$ homogénea con espacio de estados discreto. Recordemos que μ_{ij} son los tiempos esperados de primera, los cuales definimos anteriormente.

Definición 7.4. En el caso $i = j$, llamamos **tiempo esperado de recurrencia** del estado i a μ_{ii} . Observemos que μ_{ii} es el número esperado de transiciones hasta que el proceso regresa al estado inicial i .

Proposición 7.5. *Bajo las condiciones del teorema 7.1. Dada la distribución de probabilidad estacionaria $\pi = (\pi_0, \pi_1, \dots)$, se verifica que*

$$\mu_{ii} = \frac{1}{\pi_i} \text{ para } i = 0, 1, \dots$$

Demostración. Supongamos que la cadena de Markov comienza en el estado i , y sea t_1 el número de transiciones hasta que la cadena entra en el estado i ; luego sea t_2 el número adicional de transiciones desde el tiempo t_1 hasta que la cadena de Markov entra a continuación en el estado i ; luego sea t_3 el número adicional de transiciones desde el tiempo $t_1 + t_2$ hasta que la cadena de Markov entra a continuación en el estado i , y así sucesivamente.

Tenemos que tener en cuenta que t_1 es finito porque como ya vimos, si i es recurrente e i se comunica con j entonces $f_{ij} = 1$ y esto nos dice que con probabilidad 1, en algún momento ocurrirá una transición a i .

Además, para $n \geq 2$, debido a que t_n es el número de transiciones entre la $(n-1)$ -ésima y la n -ésima transición al estado i , de la propiedad de Markov se sigue que t_2, t_3, \dots son independientes e idénticamente distribuidas con media μ_{ii} .

Debido a que la n -ésima transición al estado j ocurre en el tiempo $t_1 + \dots + t_n$ obtenemos que π_i es

$$\begin{aligned} \pi_i &= \lim_{n \rightarrow \infty} \frac{n}{\sum_{j=1}^n t_j} \\ &= \lim_{n \rightarrow \infty} \frac{1}{\frac{1}{n} \sum_{j=1}^n t_j} \\ &= \lim_{n \rightarrow \infty} \frac{1}{\frac{t_1}{n} + \frac{t_2 + \dots + t_n}{n}} \\ &= \frac{1}{\mu_{ii}}, \end{aligned}$$

donde la última igualdad es debida a que $\lim_{n \rightarrow \infty} \frac{t_1}{n} = 0$ y por la ley de los grandes números $\lim_{n \rightarrow \infty} \frac{t_2 + \dots + t_n}{n} = \lim_{n \rightarrow \infty} \frac{t_2 + \dots + t_n}{n-1} \frac{n-1}{n} = \mu_{ii}$. Entonces $\mu_{ii} = \frac{1}{\pi_i}$. \square

Ejemplo 7.6. Acabamos de calcular en el ejemplo del tiempo π_0 y π_1 y entonces podemos calcular los tiempos esperados de recurrencia del estado 0 y del estado 1 utilizando la proposición anterior:

$$\mu_{00} = \frac{1}{\pi_0} = 4,$$

$$\mu_{11} = \frac{1}{\pi_1} = 1'333.$$

Para acabar, vamos a ver un ejemplo de la ley de Hardy-Weingberg, que es muy importante en biología y que representa la utilidad que pueden tener las cadenas de Markov en la vida real.

Ejemplo 7.7. Podemos encontrar una aplicación interesante de las ecuaciones del estado estable en la ley de Hardy-Weingberg, donde se aplican las cadenas de Markov a la genética.

Si consideramos una población de individuos, cada uno con un par de genes que denotamos por AA, aa ó Aa y tienen probabilidades p_0, q_0 y r_0 respectivamente, tales que $p_0 + q_0 + r_0 = 1$.

Si dos individuos se aparean, cada uno aporta uno de sus genes, elegido al azar, a la descendencia. Suponemos que cada individuo tiene la misma probabilidad de aparearse con cualquier otro individuo y lo que nos interesa es determinar con qué proporción van a aparecer los genes AA,aa y Aa en la siguiente generación, vamos a llamar a estas proporciones p, q y r .

Para empezar, vamos a tener en cuenta que elegir un progenitor al azar y luego elegir uno de sus genes es equivalente a elegir al azar un gen de la población total. Entonces obtenemos que un gen será de tipo A con probabilidad:

$$\mathbb{P}(A) = \mathbb{P}(A|AA)p_0 + \mathbb{P}(A|aa)q_0 + \mathbb{P}(A|Aa)r_0 = p_0 + \frac{r_0}{2}.$$

Y análogamente , un gen será de tipo a con probabilidad:

$$\mathbb{P}(a) = q_0 + \frac{r_0}{2}.$$

Por lo tanto, bajo apareamiento aleatorio, un miembro elegido al azar de la próxima generación será tipo AA con probabilidad p , donde:

$$p = \mathbb{P}(A)\mathbb{P}(A) = (p_0 + \frac{r_0}{2})^2.$$

de forma similar se obtiene que :

$$q = \mathbb{P}(a)\mathbb{P}(a) = (q_0 + \frac{r_0}{2})^2,$$

$$r = 2\mathbb{P}(A)\mathbb{P}(a) = 2(p_0 + \frac{r_0}{2})(q_0 + \frac{r_0}{2}).$$

Dado que cada miembro de la próxima generación será independiente de cada uno de los tres tipos de genes con probabilidades p, q, r , se deduce que los porcentajes de los miembros de la próxima generación que son del tipo AA, aa o Aa son respectivamente p, q y r .

Si consideramos el acervo genético (grupo completo de alelos únicos presentes en el material genético de la totalidad de los individuos y un alelo es cada una de las formas alternativas que puede tener un mismo gen) de esta próxima generación entonces la fracción de genes que son A ($p + \frac{r}{2}$) no cambia con respecto a la generación anterior y esto se puede ver usando lo siguiente:

$$\begin{aligned}
 p + \frac{r}{2} &= (p_0 + \frac{r_0}{2})^2 + (p_0 + \frac{r_0}{2})(q_0 + \frac{r_0}{2}) \\
 &= (p_0 + \frac{r_0}{2})[p_0 + \frac{r_0}{2} + q_0 + \frac{r_0}{2}] \\
 &= p_0 + \frac{r_0}{2} \text{ desde que } p_0 + q_0 + r_0 = 1 \\
 &= \mathbb{P}(A),
 \end{aligned} \tag{7.1}$$

o argumentando que el acervo genético total no ha cambiado de generación en generación. Por tanto, las fracciones del acervo genético que son A y a son las mismas que en la generación inicial. De esto se deduce que, bajo apareamiento aleatorio, en todas las generaciones sucesivas después de la inicial, los porcentajes de la población que tiene pares de genes AA, aa y Aa permanecerán fijos en los valores p, q y r . Esto se conoce como la ley de Hardy-Weinberg.

Supongamos ahora que la población de pares de genes se ha estabilizado en los porcentajes p, q, r y sigamos la historia genética de un solo individuo y sus descendientes (suponemos que cada individuo tiene una única descendencia, por simplicidad). Entonces, para un individuo dado, sea X_n el estado genético de su descendiente en la n -ésima generación, se tiene la siguiente matriz de transición:

$$\mathbf{P} = \begin{array}{c} \\ \\ \\ \end{array} \begin{array}{ccc} AA & aa & Aa \\ \left[\begin{array}{ccc} p + \frac{r}{2} & 0 & q + \frac{r}{2} \\ 0 & q + \frac{r}{2} & p + \frac{r}{2} \\ p + \frac{r}{4} & q + \frac{r}{4} & \frac{p}{2} + \frac{q}{2} + \frac{r}{2} \end{array} \right] \end{array}$$

Es bastante intuitivo que las probabilidades límite de esta cadena deberían ser p, q y r . Y para demostrarlo tenemos que ver que verifican las ecuaciones del estado estable, como ya

vimos que una ecuación es redundante, es suficiente comprobar que :

$$p = p\left(p + \frac{r}{2}\right) + r\left(\frac{p}{2} + \frac{r}{4}\right) = \left(p + \frac{r}{2}\right)^2,$$

$$q = q\left(q + \frac{r}{2}\right) + r\left(\frac{q}{2} + \frac{r}{4}\right) = \left(q + \frac{r}{2}\right)^2,$$

$$p + q + r = 1.$$

pero esto se sigue de la ecuación 7.1 por lo que el resultado queda probado.

Capítulo 8

Simulación de cadenas de Markov en R

Ahora que conocemos las cadenas de Markov de forma general, vamos a ver sus utilidades mediante ejemplos que podemos implementar en R.

Para empezar presentaremos un algoritmo general para simular una cadena de Markov discreta asumiendo que tenemos E estados posibles. Dada la matriz de transición \mathbf{P} , los pasos a seguir serían:

1. Establecemos el tiempo inicial $t = 0$.
2. Elegimos un estado inicial $X_t = i$.
3. Para $t = 1, \dots, T$:
 - Obtenemos la fila de \mathbf{P} correspondiente a X_t
 - Generamos X_{t+1} usando una distribución multinomial con vector de probabilidades igual a la fila que obtuvimos arriba.

Ahora implementamos esto con la siguiente función, inicializando en el estado $X_t = 1$.

```
#-----#  
#Función para simular los posibles estados de una cadena de Markov  
#-----#  
  
run.mc.sim <- function( P, num.iters = 50 ) {
```

```

#Número de posibles estados:
num.states <- nrow(P)
#Almacena los estados Xt a lo largo del tiempo
states <- numeric(num.iters)
states[1] <- 1
#Inicializa variable para el primer estado
for(t in 2:num.iters) {
#Vector de probabilidad para simular el siguiente estado Xt+1
p <- P[states[t-1], ]
#Determinar el estado
states[t] <- which(rmultinom(1, 1, p) == 1)
}
return(states)
}

```

Ejemplo 8.1. Esta función que acabamos de construir la podemos aplicar a un problema en el que conocemos la matriz de transición, usando el ejemplo del clima en Ferrol tenemos lo siguiente:

$$\mathbf{P} = \begin{bmatrix} 0'8 & 0'2 \\ 0'6 & 0'4 \end{bmatrix}$$

```

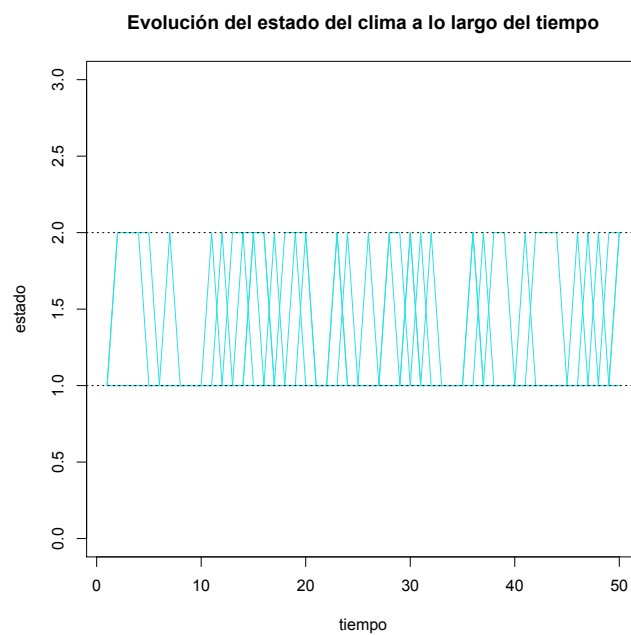
#Matriz de transición:
P <- t(matrix(c( 0.8, 0.2, 0.6, 0.4 ), nrow=2, ncol=2))
#Datos para este ejemplo:
num.chains <- 5
num.iterations <- 50
#Cada columna almacena la secuencia de estados:
chain.states <- matrix(NA, ncol=num.chains, nrow=num.iterations)
#Simulación de la cadena:
for (c in seq_len(num.chains)) {
chain.states[,c] <- run.mc.sim(P)
}

```

Nuestra función devuelve un vector que contiene los estados de nuestra cadena simulada a través del tiempo.

A continuación, visualizamos cómo evolucionan estas cadenas a lo largo del tiempo:

```
#Representación gráfica de los estados a lo largo del tiempo:
matplot(chain.states, type='l', lty=1, col=5, ylim=c(0,3), ylab='estado'
        xlab='tiempo', main="Evolución del estado del clima a lo largo
        del tiempo")
#Añadimos líneas en el 1 y en el 2:
abline(h=1, lty=3)
abline(h=2,lty=3)
```



Usando R también es fácil calcular las probabilidades del estado estable de la siguiente manera, resolviendo un sistema $Ax = b$, veámoslo:

Como vimos en el ejemplo 7.3, las ecuaciones del estado estable son

$$\pi_0 = \pi_0 0'8 + \pi_1 0'6 \implies 0'2\pi_0 - 0'6\pi_1 = 0$$

$$\pi_1 = \pi_0 0'2 + \pi_1 0'4 \implies 0'6\pi_1 - 0'2\pi_0 = 0$$

$$1 = \pi_0 + \pi_1.$$

de donde podemos sacar el sistema:

$$\underbrace{\begin{bmatrix} 0'2 & -0'6 \\ -0'2 & 0'6 \\ 1 & 1 \end{bmatrix}}_A \underbrace{\begin{bmatrix} \pi_0 \\ \pi_1 \end{bmatrix}}_x = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_b$$

Ahora usamos estos datos en R para calcular las ecuaciones del estado estable.

```
library(Matrix)
A <- matrix(c(0.2, -0.2, 1, -0.6, 0.6, 1), ncol=2,nrow=3)
b <- c(0,0,1)
#Resolvemos el sistema Ax=b:
pi <-qr.solve(A,b)
names(pi) <- c('state.1', 'state.2')
#Los resultados son los siguientes:
pi
state.1 state.2
 0.75   0.25
```

Como vemos obtenemos los mismos resultados que en el ejemplo 7.3 manualmente.

Ejemplo 8.2. Consideramos ahora el ejemplo 8.1 pero añadiendo un estado más, es decir, podemos tener un día soleado (estado 0), un día lluvioso (estado 1) o un día nublado (estado 2) y su matriz de transición es la siguiente:

$$\mathbf{P} = \begin{bmatrix} 0'3 & 0'2 & 0'5 \\ 0'6 & 0'4 & 0 \\ 0'2 & 0'7 & 0'1 \end{bmatrix}$$

Veamos la representación gráfica de este modelo, en comparación con el modelo anterior que solo tenía dos estados.

```
#Simulamos la cadena igual que en el ejemplo anterior:
P <- t(matrix(c(0.3, 0.2,0.5, 0.6, 0.4,0 , 0.2, 0.7, 0.1), nrow=3, ncol=3))
num.chains <- 5
```

```

num.iterations <- 50
chain.states <- matrix(NA, ncol=num.chains, nrow=num.iterations)
for ( c in seq_len(num.chains)) {
  chain.states[,c] <- run.mc.sim(P)
}

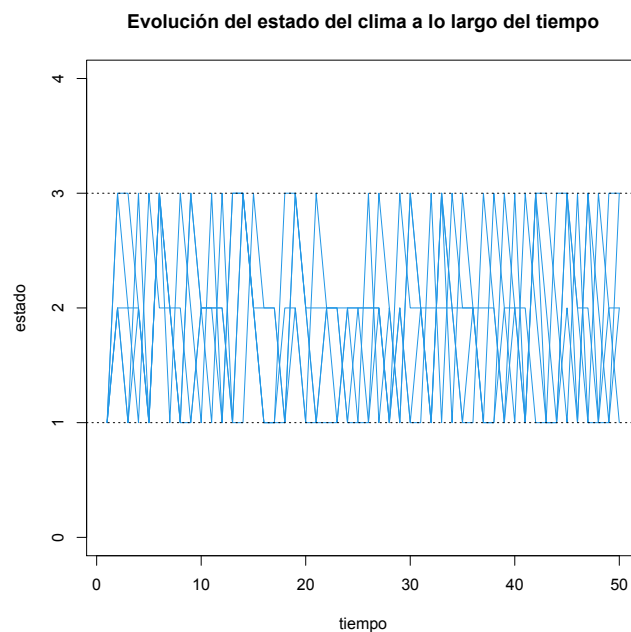
```

Con esto simulamos la cadena, igual que en el ejemplo anterior y ahora vamos a ver la representación gráfica.

```

matplot(chain.states, type='l', lty=1, col=5, ylim=c(0,3), ylab='estado'
        xlab='tiempo', main="Evolución del estado del clima a lo largo
        del tiempo")
abline(h=1, lty=3)
abline(h=3, lty=3)

```



Modelo SIR

Para finalizar vamos a hablar de los modelos SIR (utilizando la información de [10]), que se pueden modelizar con R mediante una cadena de Markov en tiempo discreto y con la ayuda del paquete "markovchain".

Estos modelos los formuló Kermack-McKendrick (1927). Sirven para estudiar epidemias de enfermedades donde la S significa susceptible de contraer la enfermedad (aún no han

contraído una enfermedad, pero no son inmunes), la I infectado (tienen la enfermedad en este momento) y la R recuperado (han tenido la enfermedad, pero ya no lo tienen y no pueden contraerlo porque se han vuelto inmunes) y son muy útiles en la vida real.

Vamos a describir como funciona el modelo, si una persona contrae la enfermedad, cambia de estado de susceptible a infectado. Si se mejoran, cambian de estado de infectado a recuperado. Mientras que es imposible cambiar de estados entre susceptible y recuperado sin pasar primero por infectado. Por otra parte es totalmente posible permanecer en el estado susceptible entre sucesivos controles de la población.

Trabajaremos con tiempo discreto ya que no tendría sentido estar controlando continuamente el estado de las personas en el sistema.

Para crear la cadena de Markov en R necesitamos lo siguiente:

- Las probabilidades de transición.
- El estado inicial.
- El paquete "markovchain".

Vamos a suponer que hay una tasa de infección del 15% y una tasa de recuperación del 25%. Eso implica que el 85% de las personas susceptibles permanecerán en el estado susceptible, y el 75% de las infectadas pasarán a recuperadas, entre pasos de tiempo sucesivos. El 100% de los recuperados permanecerán recuperados. Ninguna de las personas recuperadas se volverá susceptible.

Digamos que comienza con una población de 100 personas, donde 2 personas están infectadas. Eso significa que su `.estado inicial.es` que 98 son susceptibles, 2 están infectados y 0 están recuperados.

Así es como vamos a configurar la cadena de Markov:

```
#Cargamos el paquete que simula una cadena de Markov:
library(markovchain)
#Construyo la función que simula la cadena de Markov proporcionando los datos necesarios:
mcSIR <- new("markovchain", states=c("S","I","R"),
  transitionMatrix=matrix(data=c(0.85,0.15,0,0,0.75,0.25,0,0,1),
  byrow=TRUE, nrow=3), name="SIR")
#Doy un vector con el estado inicial
initialState <- c(98,0,2)
```

Ahora estamos en condiciones de pedir a R que nos de la matriz de transición.

```

#Datos de nuestro modelo
show(mcSIR)
SIR
A 3 - dimensional discrete Markov Chain defined by the following states:
S, I, R
The transition matrix (by rows) is defined as follows:
  S   I   R
S 0.85 0.15 0.00
I 0.00 0.75 0.25
R 0.00 0.00 1.00

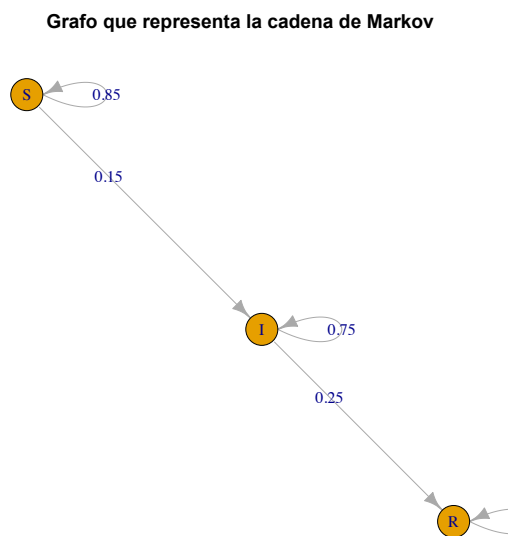
```

También podemos ver un grafo con las probabilidades de transición usando el comando.

```

#Grafo del modelo
plot(mcSIR, main="Grafo que representa la cadena de Markov")

```



Hasta ahora solo modelizamos la cadena de Markov, pero a continuación vamos a realizar una simulación que nos mostraría cuántas personas hay en cada uno de los tres estados a medida que pasa de un paso de tiempo discreto a muchos otros.

Crearemos una tabla de datos donde etiquetaremos cada paso de tiempo y un recuento de

cuántas personas hay en cada estado en cada paso de tiempo. Después llenamos esta tabla con los resultados obtenidos en cada paso de tiempo i .

```
timesteps <- 100
#Almaceno los datos en forma de hoja de datos:
sir.df <- data.frame( "timestep" = numeric(),
  "S" = numeric(), "I" = numeric(),
  "R" = numeric(), stringsAsFactors=FALSE)
for (i in 0:timesteps) {
newrow <- as.list(c(i,round(as.numeric(initialState * mcSIR ^ i),0)))
sir.df[nrow(sir.df) + 1, ] <- newrow
}
```

Ahora le pedimos a R que nos muestre los resultados calculados en la tabla

```
#Datos que obtuvimos al simular la cadena
```

```
sir.df
  timestep S I R
1         0 98 0 2
2         1 83 15 2
3         2 71 24 6
4         3 60 28 12
5         4 51 30 19
6         5 43 30 26
7         6 37 29 34
8         7 31 28 41
9         8 27 25 48
10        9 23 23 54
11       10 19 21 60
12       11 16 18 65
:
```

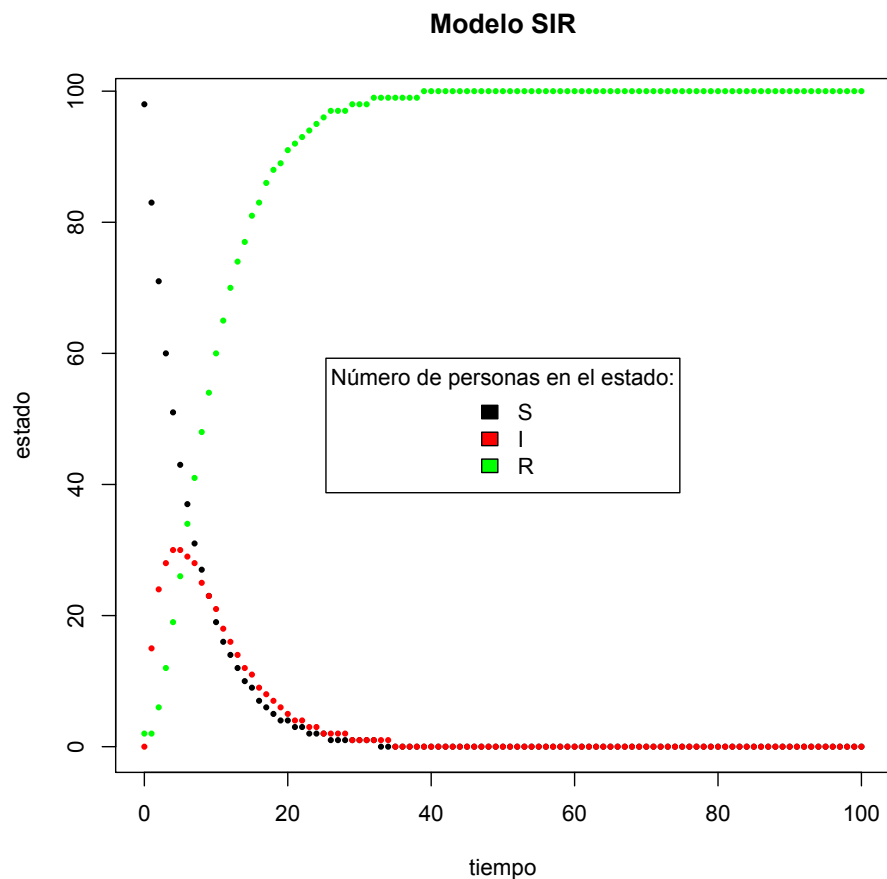
Y también lo vamos a graficar para ver los resultados:

```
#Damos las gráficas del modelo donde vemos la evolución de cada
estado a lo largo del tiempo:
plot(sir.df$timestep,sir.df$S, pch=20, cex=0.7, main= "Modelo SIR",
```

```

ylab='estado', xlab='tiempo')
legend(x="topright", legend= c("S","I","R"), fill=c("black",
      "red","green"), title="Número de personas en el estado:")
points(sir.df$timestep,sir.df$I, col="red", pch=20, cex=0.7)
points(sir.df$timestep,sir.df$R, col="green",pch=20, cex=0.7)

```



Además el paquete "markovchain" se puede utilizar para identificar elementos de su sistema a medida que evoluciona con el tiempo que pueden ser interesantes en el modelo.

```
absorbingStates(mcSIR) #Estados absorbentes
```

```
[1] "R"
```

```
transientStates(mcSIR) #Estados transitorios
```

```
[1] "S" "I"
```

```
steadyStates(mcSIR) #Estados recurrentes
```

```
  S I R
```

```
[1,] 0 0 1
```

Y estas son algunas aplicaciones en R de nuestras cadenas de Markov que podemos utilizar en la vida cotidiana entre muchas otras.

Con esto concluimos el trabajo de introducción a las cadenas de Markov donde hemos visto un estudio de las definiciones y propiedades más generales y las aplicaciones de estas cadenas en diversas áreas de estudio.

Bibliografía

- [1] Privault, N. , Springer, *Understanding Markov Chains*, 2018.
- [2] Ross, S.M. , Academic Press, *Introduction to Probability Models*, 2019.
- [3] Sericola, B. , Wiley, *Markov Chains*, 2013.
- [4] Tijms, H.C., Wiley, *A First Course in Stochastic Models*, 2003.
- [5] Dobrow, R.P. , Wiley, *Introduction to Stochastic Processes with R*, 2016.
- [6] Pinsky, M.A. and Karlin, S. , Elsevier, *An Introduction to Stochastic Modelling*, 2011.
- [7] Hillier, F. and Lieberman, G. , McGraw- Hill, *Introducción a la investigación de operaciones*, 2010.
- [8] Ross, S.M. , Wiley, *Stochastic processes*, 1996.
- [9] https://stephens999.github.io/fiveMinuteStats/index.html#Stochastic_Processes
- [10] <https://www.r-bloggers.com/2015/12/a-discrete-time-markov-chain-dtmc-sir-model-in-r/>
- [11] <https://prezi.com/yjkz-9hikpqe/tiempos-de-primera-pasada/>
- [12] <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/PEst/tema2pe.pdf>
- [13] <https://www.biografiasyvidas.com/biografia/m/markov.htm>