Matemática II - Biología

Facultad de Ciencias Exactas y Naturales - Universidad de Buenos Aires

Segundo Cuatrimestre 2020

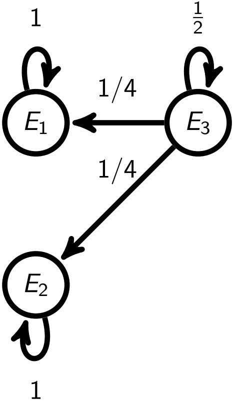
Cadenas de Markov

La clase pasda vimos que estudiando el espectro de la matriz podemos determinar si hay o no estado límite.

Sobre la existencia de estado límite

- *Si el único autovalor de P de módulo 1 es $\lambda = 1$ entonces hay estado límite para cualquier dato inicial $\mathbf{v}(0)$. O sea que todos los $\mathbf{v}(0) \in \mathcal{V}$ tienen estado límite.*
  Dicho estado límite varía según el dato inicial $\mathbf{v}(0)$ pero está en el $\varepsilon_{\lambda=1}$, o sea que es una combinación lineal de autovectores asociados a $\lambda = 1$.
- *Si además $\lambda = 1$ es un autovalor simple, i.e., $m_\lambda = 1$ entonces el estado límite $\mathbf{v}^\infty$ es único. O sea que tendremos un solo estado limite independientemente de cual sea el vector inicial.*

Ejemplo: La **migración de tres poblaciones**, está regida por el siguiente diagrama de estados:



Queremos determinar como son los estados de equilibrio y si existe estado límite

La matriz de transición es

$$P = \begin{pmatrix} 1 & 0 & 1/4 \\ 0 & 1 & 1/4 \\ 0 & 0 & 1/2 \end{pmatrix}$$

sus autovalores son $\lambda = 1$ doble y $\lambda = \frac{1}{2}$.
Autovectores:

Para $\lambda = 1$ tenemos $\mathbf{v} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ y $\mathbf{w} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ (autovectores en el espacio de estados $\mathcal{V}$)

Para $\lambda = 1/2$, tenemos $\mathbf{z} = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$ (notar que sus componnetes suman 0).

Notar que P es diagonalizable en $\mathbb{R}$. Sus autovectores son base de $\mathbb{R}^3$.

Estados de equilibrio:
Cualquier

$$\mathbf{v}^* = \begin{pmatrix} \alpha \\ 1-\alpha \\ 0 \end{pmatrix} \quad \text{con } \alpha \in [0,1]$$

es estado de equilibrio pues cumple $\mathbf{v}^* \in \mathcal{V}$ y $P\mathbf{v}^* = \mathbf{v}^*$.

Se pueden encontrar dos estados de equilibrio que sean linealmente independientes.

Por ejemplo el $(1/2, 1/2, 0)$ y $(1/3, 2/3, 0)$.
El $(1, 0, 0)$ y $(0, 1, 0)$. El $(1/4, 3/4, 0)$ y $(2/3, 1/3, 0)$.

Podemos armar infinitos pares de estados de equilibrio linealmente independientes (pero no hay 3 estados de equilibrio que sean l.i)

Como $\lambda = 1$ es el único autovalor de modulo 1 por la proposición anterior sabemos que hay estado límite.

Como $\lambda = 1$ no es simple, el estado limite no es único. Dicho estado será una combinación lineal de los autovectores asociados a $\lambda = 1$. y dependerá del vector de estados inicial.

Para el vector inicial, en miles de individuos, $\mathbf{u}(0) = (30, 20, 10)$, que podemos representar como vector en el espacio de estados como $\mathbf{v}(0) = (1/2, 1/3, 1/6)$. Buscamos el estado límite.
Escribimos el vector de estado inicial como combinación lineal de los autovectores:

$$\begin{pmatrix} 1/2 \\ 1/3 \\ 1/6 \end{pmatrix} = \frac{7}{12}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{5}{12}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \frac{1}{12}\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$$

$$\Rightarrow \mathbf{v}(n) = P^n \begin{pmatrix} 1/2 \\ 1/3 \\ 1/6 \end{pmatrix} = \frac{7}{12}1^n \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{5}{12}1^n \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \frac{1}{12}(\frac{1}{2})^n \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$$

Haciendo $n \to \infty$ concluimos que el estado limite es $\mathbf{v}^\infty = (7/12, 5/12, 0)$.
Podemos verificar que $P\mathbf{v}^\infty = \mathbf{v}^\infty$ y $\mathbf{v}^\infty = \mathbf{v}^*$ tomando $\alpha = 7/12$.

Otra forma de llegar al estado límite.
Notamos que la matriz P es diagonalizable en $\mathbb{R}$. De hecho podemos escribir

$$P = \mathrm{V} D \mathrm{V}^{-1}$$

donde

$$\mathrm{V} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 2 \end{pmatrix} \quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1/2 \end{pmatrix} \quad \mathrm{V}^{-1} = \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 1/2 \end{pmatrix}$$

Luego $P^2 = \mathrm{V} D \mathrm{V}^{-1} \mathrm{V} D \mathrm{V}^{-1} = \mathrm{V} D^2 \mathrm{V}^{-1}$, y en general

$$P^k = \mathrm{V} D^k \mathrm{V}^{-1} = \mathrm{V} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (1/2)^k \end{pmatrix} \mathrm{V}^{-1}$$

es sencillo ver como será en el límite.

$$\lim_{k\to+\infty} P^k = \mathrm{V}\,(\lim_{k\to+\infty} D^k)\,\mathrm{V}^{-1} = \mathrm{V}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\mathrm{V}^{-1}$$

$$= \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{pmatrix} = P^\infty$$

y como $\mathbf{v}(k) = P^k\mathbf{v}(0)$, podemos encontrar el estado limite usando P^∞. Para el estado inicial $\mathbf{v}(0) = (1/2, 1/3, 1/6)$ nos queda

$$\mathbf{v}^\infty = P^\infty\mathbf{v}(0) = \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1/2 \\ 1/3 \\ 1/6 \end{pmatrix} = \begin{pmatrix} 7/12 \\ 5/12 \\ 0 \end{pmatrix}$$

La conclusión es que a la larga de los 60 mil individuos tendremos 35 mil en la población 1 y 25 mil en la ciudad 2 y 0 en la ciudad 3.

Supongamos ahora que el vector inicial es $\mathbf{u}(0) = (0, 30, 30)$, que como vector de estados es $\mathbf{v}(0) = (0, 1/2, 1/2)$, y queremos predecir la distribución de la población a largo plazo.

Usando su descomposición en los autovectores de P:

$$\mathbf{v}(0) = \begin{pmatrix} 0 \\ 1/2 \\ 1/2 \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{3}{4}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$$

Nota: para hallar los coeficientes de a_1, a_2, a_3 de $\mathbf{v}(0)$ en la base $\{(1,0,0),(0,1,0),(-1,-1,2)\}$ podemos resolver el sistema

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 2 \end{pmatrix}\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1/2 \\ 1/2 \end{pmatrix}$$

o sea resolviendo el sistema $\mathrm{V}\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \mathbf{v}(0)$

Luego, como $\mathbf{v}(0) = \frac{1}{4}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{3}{4}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \frac{1}{4}\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$ y

$$P\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = 1\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad P\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = 1.\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad P\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

y $\mathbf{v}(n) = P^n\mathbf{v}(0)$ resulta

$$\mathbf{v}(n) = \frac{1}{4}1^n\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \frac{3}{4}1^n\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \frac{1}{4}(\frac{1}{2})^n\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$$

y por ende $\mathbf{v}^\infty = (1/4, 3/4, 0)$

Lo podemos buscar usando P^{∞}

$$\mathbf{v}^{\infty} = P^{\infty}\mathbf{v}(0) = \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1/2 \\ 1/2 \end{pmatrix} = \begin{pmatrix} 1/4 \\ 3/4 \\ 0 \end{pmatrix}$$

Verificamos que $P\mathbf{v}^{\infty} = \mathbf{v}^{\infty}$ y $\mathbf{v}^{\infty} = \mathbf{v}^{*}$ tomando $\alpha = 1/4$.

Esto nos dice que a la larga de los 60 mil individuos tendremos 15 mil en el estado 1 y 45 mil en el estado 2.

Nota: este es un ejemplo donde $\lambda = 1$ fue el único autovalor de módulo 1 pero **No es simple**.
Cualquiera sea el estado inicial podemos alcanzar un estado límite, pero **el estado límite depende del dato inicial.**
Lo hacemos también con Python y comparamos.

Ejemplo

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

Sus autovalores son $\lambda = 1$ y $\lambda = -1$.
Notemos que como la matriz tiene un autovalor $\lambda = -1$ que tiene modulo 1 y es $\neq 1$, entonces **no habrá estado limite** (se entiende que no hay un estado limite para cuaquier estado inicial).

De hecho, dado un estado inicial cualquiera $\mathbf{v}(0) = (a, 1-a)$, $0 \leq a \leq 1$, tendremos

$$\mathbf{v}(2k+1) = (1-a, a) \quad \text{y} \quad \mathbf{v}(2k) = (a, 1-a)$$

Los estados van a oscilar entre esos dos. La única posibilidad de convergencia es que $a = 1 - a$, o sea que tengamos $\mathbf{v}(0) = (\frac{1}{2}, \frac{1}{2})$ que es justamente un autovector asociado a $\lambda = 1$ o sea un estado de equilibrio.

Nota: este es un ejemplo donde no hay estado límite cualquiera sea el dato inicial.

Sobre la existencia de P^∞

- *Si el único autovalor de P de módulo 1 es $\lambda = 1$, o sea $\lambda = 1$ es autovalor simple o multiple y los demás autovalores cumplen $|\lambda| < 1$, entonces $\exists P^\infty$.*

  Si además $\lambda = 1$ es un autovalor simple entonces

$$P^\infty = \begin{pmatrix} \mathbf{v}^\infty & \mathbf{v}^\infty & \cdots & \mathbf{v}^\infty \\ \downarrow & \downarrow & \cdots & \downarrow \end{pmatrix}$$

- *Si existe un autovalor de P con $|\lambda| = 1$ y $\lambda \neq 1$ entonces $\nexists P^\infty$*

En el ejemplo de las tres poblaciones en que $\lambda = 1$ es autovalor doble y no teníamos otro autovalor de modulo 1

$$P^{\infty} = \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{pmatrix}$$

En el ejemplo de las flores que tenia a $\lambda = 1$ simple existe P^{∞} y es

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

Mientras que para el caso en que P era la matriz de permutación

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

cuyos autovalores eran $\lambda = 1$ y $\lambda = -1$. Tenemos

$$P^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad P^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Por lo que $P^{2k} = I$ y $P^{2k+1} = P$ y en consecuencia
No existe P^∞
Como ya sabiamos por el resultado previo pues P tiene un autovalor de módulo 1 distinto a $\lambda = 1$

Nota: P es diagonalizable y por lo tanto $P^k = V \begin{pmatrix} 1 & 0 \\ 0 & (-1)^k \end{pmatrix} V^{-1}$. Resulta claro que no existe D^∞ y por ende tampoco existe P^∞.

Otra noción es la de **matriz de Markov positiva**.

Definición: *Decimos que la matriz de Markov P es positiva si $p_{ij} > 0, \forall i, j, 1 \leq i, j \leq n$.*

Proposición

Si $P \in \mathbb{R}^{n \times n}$ es una matriz de Markov positiva entonces

- *$\lambda = 1$ es un autovalor simple.*
- *si $\lambda \neq 1$ entonces $|\lambda| < 1$*

Para las matrices positivas tenemos entonces matriz limite y un estado límite único.

Proposición

- *Sean* P, Q *dos matrices de Markov y* P *positiva* $\Longrightarrow PQ$ *es positiva.*
- *si* P^k *es positiva entonces* P^{k+s} *es positiva* $\forall s \geq 0$.

Definición: *Una matriz de Markov* P *se dice que es* **regular** *si para algún* $k \in \mathbb{N}$ *tenemos que* P^k *es positiva.*

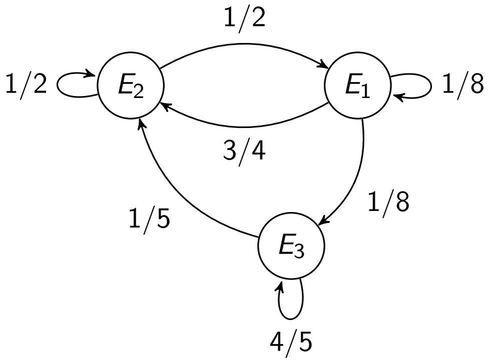
Proposición

Si $P \in \mathbb{R}^{n \times n}$ *es una matriz de Markov regular entonces*

- $\lambda = 1$ *es un autovalor simple.*
- *si* $\lambda \neq 1$ *entonces* $|\lambda| < 1$

Para las matrices regulares tenemos entonces matriz limite y un estado límite único.

Ejemplo: El movimiento entre tres estados está regida por el siguiente diagrama de estados:



Cuya matriz de transición es

$$P = \begin{pmatrix} 1/8 & 1/2 & 0 \\ 3/4 & 1/2 & 1/5 \\ 1/8 & 0 & 4/5 \end{pmatrix}$$

que No es positiva.

Sin embargo

$$P^2 = \begin{pmatrix} 25/64 & 5/16 & 1/10 \\ 79/160 & 5/8 & 13/50 \\ 37/320 & 1/16 & 16/25 \end{pmatrix}$$

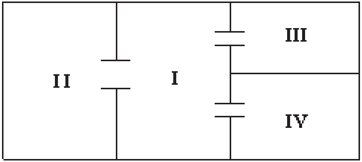
es positiva y por lo tanto P es **regular.**

Como es regular gracias a la proposición sobre matrices regulares ya sabemos que hay estado límite y que es único.

Para hallarlo solo necesitamos encontar el autovector, en el espacio de estados, asociado a $\lambda = 1$.
Hacemos la cuenta y vemos que el vector de probabilidad $\mathbf{v}^\infty$ es

$$\mathbf{v}^\infty = \begin{pmatrix} 8/27 \\ 14/27 \\ 5/27 \end{pmatrix} \text{ y } P^\infty = \begin{pmatrix} \mathbf{v}^\infty & \mathbf{v}^\infty & \mathbf{v}^\infty \\ \vdots & \vdots & \vdots \end{pmatrix} = \begin{pmatrix} 8/27 & 8/27 & 8/27 \\ 14/27 & 14/27 & 14/27 \\ 5/27 & 5/27 & 5/27 \end{pmatrix}$$

Ejemplo:
En el instante inicial 20 ratones se encuentran en la casilla I. Se supone que nada distingue un compartimento de otro, o sea que es igualmente probable que un ratón pase a cualquiera de las casillas adyacentes o se quede en la casilla en la que está. Se realizan observaciones cada hora y se registra el nuevo estado.



1. Determinar la matriz de transición P del proceso.
2. Determinar el vector de estado después de 5 horas y como es la distribución de los ratones en las casillas.
3. Mostrar un estado de equilibrio del sistema.
4. Decidir si existe o no un único estado límite.
5. Decidir si exite P^{∞} y en caso afirmativo hallarla.

$$P = \begin{pmatrix} 1/4 & 1/2 & 1/2 & 1/2 \\ 1/4 & 1/2 & 0 & 0 \\ 1/4 & 0 & 1/2 & 0 \\ 1/4 & 0 & 0 & 1/2 \end{pmatrix}$$

$$\mathbf{v}(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \mathbf{v}(5) = \begin{pmatrix} 0{,}391 \\ 0{,}203 \\ 0{,}203 \\ 0{,}203 \end{pmatrix} \quad \mathbf{v}^{\infty} = \begin{pmatrix} 2/5 \\ 1/5 \\ 1/5 \\ 1/5 \end{pmatrix}$$

$$P^{\infty} = \begin{pmatrix} 2/5 & 2/5 & 2/5 & 2/5 \\ 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 \\ 1/5 & 1/5 & 1/5 & 1/5 \end{pmatrix}$$

Distribución límite de ratones es 8 en la casilla I, y 4 en las demás casillas.

Lo hacemos también con Python y comparamos.

Un artículo reciente sobre COVID usando cadenas de Markov

A Markov Chain Model for Covid-19 Survival Analysis

Jorge Luis Romeu, Ph.D.
https://www.researchgate.net/profile/Jorge_Romeu
http://web.cortland.edu/romeu/; Email: romeu@cortland.edu

1.0 Introduction

The present *Markov Chain analysis* is intended to illustrate the power that Markov modeling techniques offer to Covid-19 studies. It is part of *our pro-bono collaboration to the American struggle against Covid-19*, whereby retired professionals would provide input, based on our long experience. You can read our *Proposal for Fighting Covid-19 and its Economic Fallout* in: https://www.researchgate.net/publication/341282217_A_Proposal_for_Fighting_Covid-19_and_its_Economic_Fallout

We have previously written *An Example of Survival Analysis Applied to Covid-19 Data*, found in https://www.researchgate.net/publication/342583500_An_Example_of_Survival_Analysis_Data_Applied_to_Covid-19, also *Multivariate Statistics in the Analysis of Covid-19 Data* and *More on Applying Multivariate Statistics to Covid-19 Data*, both of which can also be found in: https://www.researchgate.net/publication/341385856_Multivariate_Stats_PC_Discrimination_in_the_Analysis_of_Covid-19 and, as the already cited, also in our *ResearchGate* web page: https://www.researchgate.net/publication/342154667_More_on_Applying_Principal_Components_Discrimination_Analysis_to_Covid-19 These latter statistical methods provide useful tools for classification of states, regions, counties etc., according to levels of infection and other metrics.

In addition, we have written a tutorial on the use of *Design of Experiments (DOE) Applied to the Assessment Covid-19*. It provides an example of a tool for assessing and controlling appropriate levels of infection in states and regions. It can also be found in our ResearchGate web page: https://www.researchgate.net/publication/341532612_Example_of_a_DOE_Application_to_Coronavarius_Data_Analysis We have written *an evaluation of* the results of 25 years *off-shoring* tens of *thousands American jobs*, and the *impact* this has had *on US preparedness* to fight the Coronavarus Pandemic, found in: https://www.researchgate.net/publication/341685776_Off-Shoring_Taxpayers_and_the_Coronavarus_Pandemic And we have written a short study on the use of *reliability methods in the design and operation of ICU* units, that can be found in: https://www.researchgate.net/publication/342449617_Example_of_the_Design_and_Operation_of_an_ICU_using_Reliability_Principles

In this article we model the trajectory of Covid-19 infected patients into an ICU, and up to their death, using a Markov Chain. We start by considering a simple three-element state space. We then include additional states, to account for more complex situations.