Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques
Universitat de Barcelona

Cadenes de Markov i la seva Aplicació a la Neurofísica

Autor: Albert Trias Creus

Director: Dra. Olga Julià de Ferran
Realitzat a: Departament de Matemàtiques i Informàtica
Barcelona, 27 de juny de 2016

Abstract

Markov chains are primarily characterized by the probability distribution of their next state solely depending on their current one. This specific trait has enabled Markov chains to be applied in various fields of study, such as chemistry, economics and music. By expanding the theory of Markov chains to higher order Markov chains, where the probability distribution of their next state depends on their current one as well as on past ones, they also have an application in neurophysics. This work will study the theory behind Markov chains, going through higher order Markov chains and finally arriving to the expression of a powerful mathematical tool called *transfer entropy*, which quantifies the directionality and strength of connections among neurons, helping to shed some light into the mysteries of the brain.

Resum

Les cadenes de Markov es caracteritzen principalment perquè la distribució de probabilitat del seu proper estat només depèn del seu estat actual. Aquest tret principal ha permès que les cadenes de Markov tinguin aplicacions en diverses àrees, com la química, l'economia i la música. Desenvolupant la teoria de les cadenes de Markov a cadenes de Markov d'ordre superior, on la distribució de probabilitat del seu proper estat depèn tant del seu estat actual com d'altres anteriors, també tenen aplicacions en la neurofísica. Aquest treball estudiarà la teoria rere les cadenes de Markov, passant per les cadenes de Markov d'ordre superior i arribant a l'expressió d'una potent eina matemàtica anomenada *transferència d'entropia*, que quantifica la direccionalitat i força de les conexions entre neurones, col·laborant a entendre una mica millor els misteris del cervell.

Agraïments

Vull agrair a la Dra. Olga Julià la seva dedicació, paciència i ajuda al llarg d'aquest treball. També donar les gràcies als meus pares per creure en mi.

Índex

1 Introducció

El projecte

Entendre el cervell i la ment humana és, segons la pròpia Casa Blanca, un dels grans reptes del segle XXI. És un objectiu molt complicat degut a la complexitat del cervell humà pel número elevadíssim de neurones que conté, cadascuna de les quals té una quantitat important de conexions amb d'altres. L'única manera doncs de poder arribar a comprendre el funcionament del cervell amb un nivell de detall més gran és mitjançant un treball interdisciplinar. Entre aquest ampli ventall d'àrees de coneixement, les matemàtiques juguen un paper important per a resoldre aquest repte. La transferència d'entropia, una eina matemàtica, està ajudant a comprendre millor el comportament de cultius neuronals. La transferència d'entropia és una fórmula matemàtica que utilitza cadenes de Markov d'ordre superior. Així doncs, per poder entendre realment aquesta eina s'ha d'estudiar aquest tipus de cadenes que són una generalització de les cadenes de Markov. Per aquesta raó, aquest treball farà un repàs de les cadenes de Markov per anar-les entenent i a partir d'elles i de les seves propietats tenir la suficient base per construir el model més complexe de cadenes de Markov d'ordre superior. Mitjançant la transferència d'entropia, treballant amb dues neurones i prenent els seus senyals elèctrics com cadenes de Markov, es poden analitzar aquests senyals mirant més o menys estats anteriors per veure com afecten el senyal que emetrà l'altra neurona; d'aquesta manera, anar reconstruint tots els senyals dintre del cultiu i amb aquests visualitzar el comportament de les neurones dintre del cultiu.

Estructura de la Memòria

Aquest treball estarà dividit en tres grans apartats temàtics. Per una banda, el primer d'ells se centrarà en les cadenes de Markov. S'introduiran els seus trets més característics i s'estudiaran les seves propietats principals. Per a ajudar a visualitzar les cadenes de Markov s'empreran dos exemples, el del passeig aleatori unidimensional i el procés discret al fer una cua, que seran recurrents en aquest primer apartat. En el segon apartat s'extendrà la teoria del primer i es tractaran les cadenes de Markov d'ordre superior. Per tal d'estudiar aquest tipus de cadenes de Markov, es presentarà un model amb les seves propietats i al final es farà servir un exemple per explicar el procediment del model. Ja per últim, en el darrer apartat es tractarà directament l'aplicabiltat de les cadenes de Markov en la neurofísica i s'introduirà el concepte de *transferència d'entropia*, i al final es veurà, mitjançant un exemple real, com s'apliquen aquests coneixement en la neurofísica per extreure informació sobre les conexions entre neurones.

2 Definicions Introductòries

Comencem introduint els conceptes bàsics de probabilitat per tal de poder definir les cadenes de Markov.

Definició 2.0.1. Una família de subconjunts $\mathcal{F}$ d'Ω és una **σ-àlgebra** a Ω si satisfà les següents condicions:

1. $\emptyset \in \mathcal{F}$
2. $\forall \text{ F} \in \mathcal{F} \Rightarrow \text{F}^C \in \mathcal{F}$
3. $\forall \{\text{F}_i\}_1^\infty \subset \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} \text{F}_i \in \mathcal{F}$

Definició 2.0.2. La σ-àlgebra dels **conjunts de Borel** $\mathbb{B}$ és la σ-àlgebra en $\mathbb{R}$ generada, per exemple, pels oberts.

Definició 2.0.3. El triplet $(\Omega, \mathcal{F}, \mathbb{P})$ s'anomena **espai de probabilitat**, on:

1. el conjunt Ω és l'espai mostral i els seus elements ω s'anomenen resultats.
2. $\mathcal{F}$ és una σ-àlgebra dels subconjunts d'Ω.
3. la probabilitat $\mathbb{P} : \mathcal{F} \to [0, 1]$ és una funció satisfent que
   (a) $\mathbb{P}\{\Omega\} = 1$
   (b) σ-additivitat: $\forall \{A_i\} \subset \mathcal{F}$ disjunts, és a dir, $A_i \cap A_j = \emptyset \; \forall i \neq j$, es compleix que $\mathbb{P}\Big\{\bigcup_{i=1}^{\infty} A_i\Big\} = \sum_{i=1}^{\infty} \mathbb{P}\{A_i\}$

Definició 2.0.4. Sigui $(\Omega, \mathcal{F}, \mathbb{P})$ un espai de probabilitat. Es diu que l'aplicació $X : \Omega \to \mathbb{R}$ és una **variable aleatòria** si satisfà que

$$\forall B \in \mathbb{B}(\mathbb{R}), X^{-1}(B) \in \mathcal{F}$$

Definició 2.0.5. Un **vector aleatori** m-dimensional $X = (x_1, x_2, ..., x_m)$ és una aplicació $X : \Omega \to \mathbb{R}^m$ tal que cadascun dels seus components és una variable aleatòria.

Definició 2.0.6. $\forall A, B \subset \mathcal{F}$, amb $\mathbb{P}\{B\} > 0$, es defineix la **probabilitat condicionada** $\mathbb{P}\{A|B\}$ de l'esdeveniment A donat l'esdeveniment B com

$$\mathbb{P}\{A|B\} = \frac{\mathbb{P}\{A \cap B\}}{\mathbb{P}\{B\}}$$

Definició 2.0.7. Un **procés estocàstic** és una col·lecció de variables aleatòries $\{X_t, t \in T\}$ definides en el mateix espai de probabilitat $(\Omega, \mathcal{F}, \mathbb{P})$. El conjunt T s'anomena el conjunt de paràmetres.

Definició 2.0.8. Una **matriu estocàstica** és aquella matriu d'elements majors o iguals a 0 que compleix que la suma per columnes dels seus elements és igual a 1.

3 Cadenes de Markov

3.1 Definicions i Propietats Bàsiques

Definició 3.1.1. Sigui $\{X_n\}$ un procés on el seu conjunt d'estats I, és a dir el conjunt de valors que pot prendre el procés, és un conjunt finit o numerable. Es diu que aquest procés és una **cadena de Markov** si compleix la *propietat de Markov*, és a dir,

$$\mathbb{P}\{X_n = i_n | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}\} = \mathbb{P}\{X_n = i_n | X_{n-1} = i_{n-1}\}, \ \forall i_0, \dots, i_n \in I$$

Observació 3.1.1. Notem que $X_n = i$ expressa que X_n es troba en l'estat i.

De la definició 3.1.1. veiem que un procés de Markov està caracteritzat per ser un procés *sense memòria*, és a dir, que només intervé l'estat en que es trobava el procés justament en l'instant anterior i no la seqüència d'esdeveniments passats.

Definició 3.1.2. La **probabilitat de transició en un pas**, és a dir, la probabilitat que X_{n+1} es trobi en l'estat j donat que X_n es troba en l'estat i, es denota per $P_{ij}^{n,n+1}$

$$P_{ij}^{n,n+1} = \mathbb{P}\{X_{n+1} = j | X_n = i\}$$

Observació 3.1.2. Notem que en la definició anterior els superíndexs indiquen que les probabilitats de transició també depenen del temps de transició.

Definició 3.1.3. Un procés de Markov es diu que té unes **probabilitats de transició estacionàries** si les seves probabilitats de transició en un pas són independents del temps.

A partir d'ara podem assumir que les cadenes de Markov amb les que treballem tenen probabilitats de transició estacionàries a menys que s'indiqui el contrari. Per tant, amb aquesta suposició, tenim que $P_{ij}^{n,n+1} = P_{ij}$.

Aquests valors P_{ij}, que representen la probabilitat de passar de l'estat i al j en un pas, els col·locarem en la matriu $\mathbf{P} = ||P_{ij}||$.

Definició 3.1.4. Aquesta matriu $\mathbf{P}$ s'anomena la **matriu de Markov** o bé la **matriu de probabilitat de transició** del procés:

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ \vdots & \vdots & \vdots & \\ P_{i0} & P_{i1} & P_{i2} & \dots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

Observació 3.1.3. Notem que una matriu de Markov pot tenir dimensió infinita. En cas però que tingui dimensió finita, tindrem una matriu quadrada $\mathbf{P}$ de $m \times m$, on m serà el nombre d'estats disponibles en el procés.

Propietats 3.1.1. Veiem que els elements de la matriu de Markov tenen dues propietats principals:

1. $0 \leq P_{ij} \leq 1 \ \forall i, j \in I$
2. $\sum_{j \in I} P_{ij} = 1 \ \forall i \in I$

Veiem que la propietat 1 es compleix ja que cada element P_{ij} és una probabilitat. Per altra banda, la propietat 2 és certa perquè deixant fix l'estat de sortida i el procés haurà d'anar a algun estat j i la suma de probabilitats de tots els possibles estats j als que pot anar ha de ser total, és a dir, 1.

Observació 3.1.4. Notem que degut al segon punt de la propietat anterior tota matriu de Markov és una matriu estocàstica.

Teorema 3.1.1. Sigui $\{X_n\}$ una cadena de Markov, λ la probabilitat de distribució de X_0 i $\mathbf{P}$ la seva matriu de Markov. Aleshores la llei de $\{X_n\}$ queda totalment determinada per λ i $\mathbf{P}$ i diem que $\{X_n\}$ és Markov(λ, $\mathbf{P}$).

Demostració 3.1.1. Per demostrar el teorema procedirem a expressar, mitjançant només λ i P_{ij}, la probabilitat $\mathbb{P}\{X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n\}$.

Per la definició de probabilitat condicionada tenim que

$$\mathbb{P}\{X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n\} = \mathbb{P}\{X_n = i_n | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\} \cdot$$

$$\cdot \mathbb{P}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\}$$

Per definició d'un procés de Markov tenim que

$$\mathbb{P}\{X_n = i_n | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\} = \mathbb{P}\{X_n = i_n | X_{n-1} = i_{n-1}\} = P_{i_{n-1}, i_n}$$

Per tant, tindrem que

$$\mathbb{P}\{X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n\} = P_{i_{n-1}, i_n} \cdot \mathbb{P}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\}$$

on, seguint el mateix raonament,

$$\mathbb{P}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}\} = P_{i_{n-2}, i_{n-1}} \cdot \mathbb{P}\{X_0 = i_0, X_1 = i_1, \ldots, X_{n-2} = i_{n-2}\}$$

Així doncs, iterant aquest mateix procés, al final obtindrem,

$$\mathbb{P}\{X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_n = i_n\} = P_{i_{n-1}, i_n} P_{i_{n-2}, i_{n-1}} \ldots P_{i_0, i_1} \cdot \mathbb{P}\{X_0 = i_0\} =$$

$$= P_{i_{n-1}, i_n} P_{i_{n-2}, i_{n-1}} \ldots P_{i_0, i_1} \lambda_{i_0} \quad \square$$

Definició 3.1.5. La **probabilitat de transició en n passos**, $n \geq 0$, és la probabilitat que X_{n+m} es trobi en l'estat j donat que X_m es trobi en l'estat i, i es denota per P_{ij}^n

$$P_{ij}^n = \mathbb{P}\{X_{n+m} = j | X_m = i\}$$

Observació 3.1.5. En el cas $n = 0$ aquesta probabilitat de transició es defineix de la següent manera:

$$P_{ij}^0 = \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Per altra banda, també hi haurà una **matriu de probabilitat de transició de n-passos** que es denota per $\mathbf{P}^{(n)} = ||P_{ij}^n||$

Teorema 3.1.2. Sigui $\{X_n\}$ Markov $(\lambda, \mathbf{P})$ i sigui $\mathbf{P}^{(n)}$ la seva matriu de probabilitat de transició de n-passos. Aleshores es compleix que $\mathbf{P}^{(n)} = \mathbf{P}^n$.

Demostració 3.1.2. Sigui I el conjunt d'estats possibles que pot prendre $\{X_n\}$.

En el cas en que $\mathbb{P}(X_n = i) > 0$ es té:

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

$$\mathbb{P}(X_{n+2} = j | X_n = i) = \sum_{k \in I} \mathbb{P}(X_{n+2} = j, X_{n+1} = k | X_n = i) \stackrel{\text{a}}{=} \sum_{k \in I} \mathbb{P}(X_{n+1} = k | X_n = i) \cdot$$

$$\cdot \mathbb{P}(X_{n+2} = j | X_{n+1} = k, X_n = i) \stackrel{\text{b}}{=} \sum_{k \in I} \mathbb{P}(X_{n+1} = k | X_n = i) \cdot \mathbb{P}(X_{n+2} = j | X_{n+1} = k) =$$

$$= \sum_{k \in I} P_{ik} P_{kj} \stackrel{\text{c}}{=} \mathbf{P}_{ij}^{(2)}$$

Notem que la probabilitat $\mathbb{P}(X_{n+2} = j | X_{n+1} = k, X_n = i)$ no sempre està ben definida però en aquest cas no és un problema. Si $\mathbb{P}(X_{n+1} = k, X_n = i) > 0$, sí que estarà ben definida. En cas que $\mathbb{P}(X_{n+1} = k, X_n = i) = 0$ es tindrà que $\mathbb{P}(X_{n+1} = k | X_n = i) = 0$ i es compleix que $\mathbb{P}(X_{n+1} = k | X_n = i) \cdot \mathbb{P}(X_{n+2} = j | X_{n+1} = k, X_n = i) = 0$ encara que $\mathbb{P}(X_{n+2} = j | X_{n+1} = k, X_n = i)$ no estigui definida. Per tant, a la igualtat a s'ha pogut utilitzar la probabilitat condicionada. A la igualtat b s'ha utilitzat la propietat de Markov.

Ja per últim, a la igualat c veiem que $\sum_{k \in I} P_{ik} P_{kj}$ és l'element ij de la matriu $\mathbf{P}^2$. Per tant, arribem a que $\mathbf{P}_{ij}^2 = \mathbf{P}_{ij}^{(2)}$

A continuació procedim a demostrar per inducció el cas general, és a dir, que:

si $\mathbb{P}(X_n = i) > 0$, aleshores $\mathbb{P}(X_m = j | X_n = i) = P_{ij}^{(m-n)} = P_{ij}^{m-n}$, $\forall m > n$

Suposem que és cert pel cas $m - 1 > n$ i veiem que també és cert pel cas $m > n$.

$$\mathbb{P}(X_m = j | X_n = i) = \sum_{k \in I} \mathbb{P}(X_m = j, X_{m-1} = k | X_n = i) = \sum_{k \in I} \mathbb{P}(X_{m-1} = k | X_n = i) \cdot$$

$$\cdot \mathbb{P}(X_m = j | (X_{m-1} = k, X_n = i)) = \sum_{k \in I} \mathbb{P}(X_{m-1} = k | X_n = i) \cdot \mathbb{P}(X_m = j | X_{m-1} = k) =$$

$$= \sum_{k \in I} P_{ik}^{(m-n-1)} P_{kj} \stackrel{\text{d}}{=} \sum_{k \in I} P_{ik}^{m-n-1} P_{kj} = P_{ij}^{m-n}$$

A la igualtat d utilitzem la hipòtesi d'inducció.

Per tant queda demostrat que $\mathbf{P}^{(n)} = \mathbf{P}^n$ $\square$

3.2 Exemples

A continuació introduïrem dos exemples de cadenes de Markov als que ens referirem al llarg del treball per tal de comprendre millor algunes de les seves propietats més importants.

Exemple 3.2.1. ***Passeig aleatori unidimensional***

Imaginem una persona que només pot caminar de forma unidimensional sobre una recta però de forma aleatòria i cada cop que fa una passa es comptabilitza la seva posició. Tenim que aquesta situació es pot descriure per una cadena de Markov on el seu conjunt d'estats serà un subconjunt finit o infinit dels enters. Per fer la representació de la matriu de Markov d'aquest procés més fàcil, el conjunt d'estats I serà els nombres naturals. Així doncs, en cas que s'arribi a l'estat 0, no es podrà accedir a estats negatius sinó que el caminant es quedarà en l'estat 0 o passarà a estats positius.

Suposem que aquesta persona es troba en l'estat o posició i. Sigui p_i la probabilitat que faci una passa cap endavant, i per tant estigui en la posició $i+1$, i q_i la que faci una passa cap enrere i es trobi en la posició $i-1$. Per fer la situació encara més general, sigui r_i la probabilitat que la persona no faci cap passa i romangui en la posició i.

Formalment aquestes probabiliats es descriuran de la forma següent:

$\mathbb{P}\{X_{n+1} = i+1 | X_n = i\} = p_i \ \forall n \geq 1 \ \forall i \in I$

$\mathbb{P}\{X_{n+1} = i-1 | X_n = i\} = q_i \ \forall n \geq 1 \ \forall i \in I \setminus \{0\}$

$\mathbb{P}\{X_{n+1} = i | X_n = i\} = r_i \ \forall n \geq 1 \ \forall i \in I$

Tal com està plantejat aquest exemple i al ser p_i, q_i i r_i probabilitats se satisfarà que:

a) $0 \leq q_i \leq 1, \ 0 \leq r_i \leq 1, \ 0 \leq p_i \leq 1, \ \text{i} \ q_i + r_i + p_i = 1 \quad \forall i \in I \setminus \{0\}$

b) $0 \leq r_0 \leq 1, \ 0 \leq p_0 \leq 1, \ \text{i} \ r_0 + p_0 = 1$

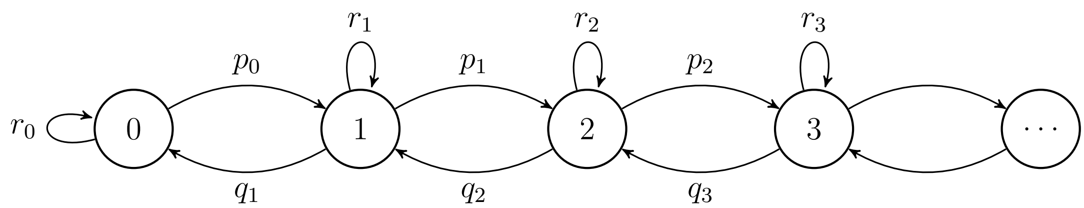

Figura 1: *Representació del comportament de la cadena de Markov pel cas del passeig aleatori unidimensional.*

A la Figura 1 es pot veure que no intervé q_0 perquè tal com està definit I, quan la persona es troba en l'estat 0 no està permès que faci una passa cap enrere.

Així doncs, la matriu de Markov d'una passeig aleatori unidimensional amb les restriccions explicades aquí serà la següent:

$$\mathbf{P} = \begin{bmatrix} r_0 & p_0 & 0 & 0 & 0 & \dots \\ q_1 & r_1 & p_1 & 0 & 0 & \dots \\ 0 & q_2 & r_2 & p_2 & 0 & \dots \\ 0 & 0 & q_3 & r_3 & p_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Exemple 3.2.2. ***Procés discret al fer una cua***

Suposem que ens trobem en una botiga on van arribant els clients i fan cua per ser atesos i que només s'atèn a un client per torn. En cas que no hi hagi ningú fent cua no s'atendrà a cap client. Durant un torn poden arribar nous clients a la botiga. Sigui el nombre de nous clients en el torn n-èssim la variable aleatòria ξ_n amb una funció de distribució independent del torn. Assumim que les variables aleatòries ξ_n són independents i idènticament distribuïdes amb probabilitats

$$\mathbb{P}\{\xi_n = k\} = a_k, \quad \text{on } \ 0 \le a_k \le 1 \ \ \forall k \in \mathbb{N} \ \text{ i } \ \sum_{k=0}^{\infty} a_k = 1$$

Notem que l'estat del sistema a l'inici de cada torn ve donat per quants clients hi hagi fent cua. Així doncs, si en el torn n ens trobem a l'estat i i ξ_n representa els clients nous que han entrat durant el torn n-èssim, al següent torn estarem en l'estat j amb valor

$$j = \begin{cases} i - 1 + \xi & \text{si } i \ge 1 \\ \xi & \text{si } i = 0 \end{cases}$$

El procés es pot descriure també de la forma:

$$X_{n+1} = \max(X_n - 1, 0) + \xi_n$$

Amb tot això, veiem que en passar de l'estat i del sistema al seu següent, j, el valor de ξ pot ser indefinidament gran i per tant sempre és possible arribar a un valor de j major. En canvi, només hi ha un cas en el que el valor de j serà inferior al de i: quan en el torn teníem com a mínim un client, l'hem atès i no n'ha entrat cap de nou. En aquest cas el valor de j serà de $i-1$. Hem vist doncs que no és possible en aquesta situació arribar a un valor j que tingui un valor inferior a $i - 1$ en un sol pas. Així doncs, la matriu de Markov d'aquest procés només tindrà elements P_{ij} no nuls si es compleix que $j \ge \max(i - 1, 0)$.

Com hem definit $\mathbb{P}\{\xi_n = k\} = a_k$, la matriu de Markov $\mathbf{P}$ d'aquest procés serà de la forma següent:

$$\mathbf{P} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

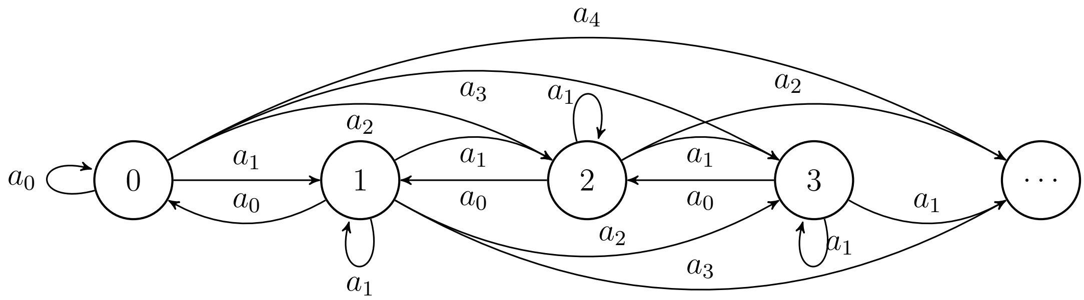

Figura 2: *Representació del comportament de la cadena de Markov pel cas del procés discret al fer una cua.*

3.3 Classificació d'Estats

A continuació procedim a estudiar els subcojunts d'estats anomenats classes de les cadenes de Markov, que fragmenten la cadena en unitats més petites que faciliten després poder entendre la cadena en la seva totalitat.

Definició 3.3.1. Siguin i i j dos estats possibles de la cadena de Markov $\{X_n\}$. Es diu que i **condueix** a j o que j és **accessible** des de l'estat i i s'expressa com $i \to j$ si

$$\exists\ n > 0 \mid P^n_{ij} > 0$$

és a dir, $\exists\ n$ tal que és possible anar de i a j en n-passos.

Definició 3.3.2. En el cas en que i condueix a j i que també j condueix a i es diu que els estats i i j es **comuniquen**. Això s'expressa per $i \leftrightarrow j$.

Observació 3.3.1. Dos estats i i j no es comunicaran només en els dos casos següents:

1. $P^n_{ij} = 0\ \forall n \geq 0$
2. $P^n_{ji} = 0\ \forall n \geq 0$

Clarament tampoc es comunicaran si ambdós casos són certs.

Proposició 3.3.1. La relació de comunicació $\leftrightarrow$ és una relació d'equivalència sobre el conjunt d'estats I. Això permet fer particions de I en classes d'equivalència. Com en aquestes classes d'equivalència tots els estats comuniquen entre si, s'anomenaran **classes de comunicació**.

Demostració 3.3.1. Demostrem que la relació $\leftrightarrow$ és reflexiva, simètrica i transitiva.

i. *Reflexiva*

Prenent $n = 0$, es té que $P_{ij}^0 = \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$

ii. *Simètrica*

Si $i \leftrightarrow j \iff \begin{cases} i \rightarrow j \\ j \rightarrow i \end{cases} \iff j \leftrightarrow i$

iii. *Transitiva*

Si $i \leftrightarrow j$ i $j \leftrightarrow k$ aleshores es compleix que $\exists\ a, b \geq 0 \mid P_{ij}^a > 0$ i $P_{jk}^b > 0$.

Alhora es compleix que $P_{ik}^{a+b} = \sum_{l=0}^{\infty} P_{il}^a P_{lk}^b \geq P_{ij}^a P_{jk}^b > 0$

Amb un procediment totalment equivalent s'arriba a que $\exists\ \alpha > 0 \mid P_{ki}^{\alpha} > 0$

Per tant, $i \leftrightarrow k$ $\square$

Definició 3.3.3. Sigui C una classe de comunicació. Es diu que C és **tancada** si

$$\forall\ i \in C,\ i \rightarrow j\ \Rightarrow\ j\ \in\ C$$

D'aquesta definició s'aprecia que un cop s'arriba a un estat d'una classe tancada no hi ha manera d'escapar d'aquesta classe.

Definició 3.3.4. L'estat i s'anomena **absorbent** si $\{i\}$ és una classe tancada.

A la proposició 3.3.1. ja s'ha esmentat com la relació $\leftrightarrow$ permet fer particions d'I en classes de comunicació i que a més a més aquestes seran classes d'equivalència. Per tant, dintre de la mateixa classe d'equivalència es trobaran els estats que es comuniquen entre si.

Sigui A una classe i $i \in A$. Si A no és una classe tancada, hi ha una probabilitat positiva que $i \rightarrow j$ on $j \notin A$. En cas que això passés, seria impossible que $j \rightarrow i$ perquè sinó els estats i i j comunicarien i hi hauria una contradicció ja que $j \notin A$. Així doncs és important notar que sí que és possible passar d'un estat a un altre encara que estiguin en classes diferents.

Definició 3.3.5. Sigui $\{X_n\}$ una cadena de Markov $(\lambda, \mathbf{P})$. Si el seu conjunt d'estats I es redueix a una sola classe de comunicació, la cadena de Markov s'anomena **irreductible**.

Observació 3.3.2. Cal notar que de la definició anterior s'extreu que en una cadena de Markov irreductible tots els seus estats comuniquen entre sí.

Definició 3.3.6. Sigui $i \in I$. El **període** de l'estat i, $d(i)$, es defineix com el màxim comú divisor, (m.c.d), de tots els $n \geq 1$ tals que compleixen $P^n_{ii} > 0$.

Proposició 3.3.2. Siguin $i, j \in I$. Si $i \leftrightarrow j \Rightarrow d(i) = d(j)$.

Amb aquesta proposició queda palès que el període és una constant en cada classe de comunicació.

Definició 3.3.7. Una cadena de Markov on tots els seus estats tenen període 1 s'anomena una cadena **aperiòdica**

A continuació utilitzem els exemples de l'apartat 3.2 per representar el conceptes anteriors.

Exemple 3.3.1. ***Passeig aleatori unidimensional***

Per fer l'exemple més manipulable, prendrem el cas on és equiprobable fer una passa enrere, mantenir-se quiet, o fer una passa cap endavant. També prendrem que hi ha n estats a I i que la persona té com a meta el darrer estat $n-1$ i per tant un cop hi arribi s'hi voldrà quedar. Amb aquests condicionants la matriu de Markov del procés de dimensió $n \times n$ queda de la següent forma:

$$\mathbf{P} = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ n{-}2 \\ n{-}1 \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & \dots & n{-}1 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \dots & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\ 0 & \dots & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \dots & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Mirant doncs la matriu $\mathbf{P}$ es veu que I es redueix a les dues classes de comunicació següents: $\{0, 1, 2, \dots, n-2\}$ i $\{n-1\}$. Els estats de la primera classe es comuniquen entre si ja que la persona sempre pot caminar o cap endavant o cap enrere, a excepció de l'estat 0 on no es pot anar cap enrere per construcció i l'estat $n-2$ que només comunicarà amb els demés si no va cap endavant. Si ho fes, s'entraria a l'estat $n-1$ i aquest estat no comunica amb la resta d'estats. Recalquem que tot i que qualsevol estat de la primera classe pot acabar a la segona, veiem que des d'aquesta mai podrà retornar a la seva classe inicial. Això es deu a que la classe $\{n-1\}$ és tancada. Per tant, com hi ha dues classes de comunicació diferents, queda palès que aquesta cadena de Markov no és irreductible. Per altra banda veiem que la diagonal de la matriu mai s'anul·la, és a dir, $P_{ii} > 0 \; \forall \; i \in I$. Així doncs, veiem que $d(i) = 1, \forall i \in I$, ja que per cada estat és possible tornar a ell mateix en un sol pas i al calcular el m.c.d sempre sortirà 1. Per tant, aquesta cadena de Markov és aperiòdica.

Exemple 3.3.2. *Procés discret al fer una cua*

Recordem la matriu de Markov d'aquest procés:

$$\mathbf{P} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

Primer estudiem les classes de comunicació d'aquest procés. Tal i com està descrit, no es pot disminuir el nombre de clients en més d'un per cada torn. Així doncs, l'única possiblitat de perdre clients és quan se n'atèn a un i no n'entra cap de nou. Aquesta situació ve descrita per la probabilitat a_0. Per tant, si es vol que $i \to j$ comuniquin, on $j < i$ és necessari que $a_0 > 0$ ja que sinó no comunicarien. Com sempre, en cas que ocorri, es disminuirà el nombre de clients un per un. Mentre que $\exists m \in \mathbb{Z}^+ | a_m > 0$, aleshores tots els estats seran accesibles i estaran comunicats entre sí.

Per tant, aquesta cadena de Markov serà irreductible si es compleix que

$$0 < a_0 < 1$$

Suposem que la cadena de Markov satisfà que $0 < a_0 < 1$ i aleshores com a mínim hi ha una altra probabilitat a_m que no s'anul·la. Així doncs, sempre se satisfarà que la cadena serà aperiòdica, és a dir, $d(i) = 1 \ \forall i \in I$. Per veure-ho, suposem que ens trobem en l'estat i. Com es compleix que $0 < a_0 < 1$, és possible arribar a l'estat 0 començant des de l'i. Suposem un camí K que retorna a l'estat i en p passes havent passat per l'estat 0. Ens adonem que aquest mateix camí K es pot modificar per tal que es retorni a l'estat i en $p + 1$ passes. Això és possible, ja que com el camí K passa per l'estat 0, i degut a que $0 < a_0 < 1$, és possible passar de l'estat 0 a l'estat 0. Així doncs, veiem que hi ha dos camins possibles que comencen en l'estat i i hi retornen en p i $p + 1$ passes. Com el m.c.d(p, $p + 1$) = 1, veiem que efectivament $d(i) = 1 \ \forall i \in I$. Per tant, veiem que aquesta cadena de Markov és aperiòdica.

3.4 Recurrència

Sigui $i \in I$ un estat de la cadena $\{X_n\}$ de Markov $(\lambda, \mathbf{P})$. De manera genèrica s'entèn que l'estat i és *recurrent* si és un estat al qual s'hi retorna al llarg del procés amb probabilitat 1. En canvi, s'entèn que l'estat i és *transitori* si la probabilitat de no retornar-hi mai més és positiva. L'estat i només podrà ser o recurrent o transitori. A continuació s'estudiarà aquest fenomen amb més profunditat.

3.4.1 Definicions

Definició 3.4.1. Sigui $i \in I$. Sigui f_{ii}^n la probabilitat que trobant-se inicialment en l'estat i el primer cop que el procés retorni a aquest mateix estat sigui en la n-èssima transició:

$$f_{ii}^n = \mathbb{P}\{X_n = i, X_\mu \neq i, \mu = 1, 2, \ldots, n-1 | X_0 = i\}$$

Observació 3.4.1. Veiem dos casos particulars de f_{ii}^n:

i) per convenció, es té que $f_{ii}^0 = 0 \quad \forall i \in I$

ii) per definició, es té que $f_{ii}^1 = P_{ii} \quad \forall i \in I$

Definició 3.4.2. Siguin $i, j \in I$. Sigui f_{ij}^n la probabilitat que trobant-se inicialment en l'estat i el primer cop que el procés es trobi en l'estat j sigui en la n-èssima transició:

$$f_{ij}^n = \mathbb{P}\{X_n = j, X_\mu \neq j, \mu = 1, 2, \ldots, n-1 | X_0 = i\}$$

Teorema 3.4.1. Sigui $i \in I$, aleshores:

$$P_{ii}^n = \sum_{k=0}^{n} f_{ii}^k P_{ii}^{n-k}, \quad n \geq 1$$

Cal esmentar que aquest teorema serveix per calcular de manera recursiva els diferents valors de f_{ii}^n. Procedim a demostrar-lo.

Demostració 3.4.1. Sigui E_k l'esdeveniment que descriu que el procés comença en l'estat i i hi retorna després de n transicions amb la condició que el primer cop que retorni a l'estat i sigui en la transició k-èssima. Per tant, E_k s'expressa com $E_k = \{X_0 = i, X_1 \neq i, \ldots, X_{k-1} \neq i, X_k = i, X_n = i\}$ i la seva probabilitat és:

$$\mathbb{P}\{E_k\} = f_{ii}^k \cdot \mathbb{P}\{X_n = i | X_k = i\} = f_{ii}^k P_{ii}^{n-k}, \quad 1 \leq k \leq n$$

Per descriure el conjunt $\{X_n = i\}$ s'han de tenir en compte els diferents valors de k, i per tant,

$$P_{ii}^n = \mathbb{P}\{X_n = i | X_0 = i\} = \sum_{k=1}^{n} \mathbb{P}\{E_k\} = \sum_{k=1}^{n} f_{ii}^k P_{ii}^{n-k} \overset{\text{a}}{=} \sum_{k=0}^{n} f_{ii}^k P_{ii}^{n-k}$$

Cal notar que a la igualtat a s'ha utilitzat el punt i) de l'observació 3.4.1. □

Definició 3.4.3. L'estat $i \in I$ es **recurrent** si i només si $\displaystyle\sum_{n=1}^{\infty} f_{ii}^n = 1$

Definició 3.4.4. Si l'estat i no és recurrent s'anomena **transitori**

Teorema 3.4.2. Un estat i és recurrent si i només si $\sum_{n=1}^{\infty} P_{ii}^n = \infty$.

Proposició 3.4.1. Sigui R el número de retorns a l'estat i d'una cadena de Markov. Aleshores, $E[R|X_0 = i] = \sum_{n=1}^{\infty} P_{ii}^n$

Demostració 3.4.2. Per començar veiem que R es pot expressar mitjançant indicadors de la forma següent: $R = \sum_{n=1}^{\infty} \mathbb{1}_n$ on $\mathbb{1}_n = \mathbb{1}_{\{X_n = i\}}$, $n = 1, 2, \dots$ Per tant,

$$E[R|X_0 = i] = E\left[\sum_{n=1}^{\infty} \mathbb{1}_n | X_0 = i\right] = \sum_{n=1}^{\infty} \mathbb{P}\{X_n = i | X_0 = i\} = \sum_{n=1}^{\infty} P_{ii}^n \ \square$$

Mitjançant el teorema i la proposició anteriors, també es pot caracteritzar un estat i com a **recurrent** si i només si el número esperat de retorns a ell mateix és infinit.

Proposició 3.4.2. Si $i \leftrightarrow j$ i i és un estat recurrent, aleshores l'estat j també serà recurrent.

Demostració 3.4.3. Com que $i \leftrightarrow j$, aleshores $\exists\ a, b \geq 1$ tal que $P_{ij}^a > 0$ i $P_{ji}^b > 0$.

Prenent $\mu > 0$ se satisfà que $P_{jj}^{a+b+\mu} \geq P_{ji}^a P_{ii}^\mu P_{ij}^b$ i per tant,

$$\sum_{n=0}^{\infty} P_{jj}^n \geq \sum_{\mu=0}^{\infty} P_{jj}^{a+b+\mu} \geq \sum_{\mu=0}^{\infty} P_{ji}^a P_{ii}^\mu P_{ij}^b = P_{ji}^a P_{ij}^b \sum_{\mu=0}^{\infty} P_{ii}^\mu$$

Com es té que i és un estat recurrent, es compleix que $\sum_{\mu=0}^{\infty} P_{ii}^\mu$ divergeix i en conseqüència també divergirà per l'estat j $\square$.

Aquesta proposició il·lustra que la recurrència és una propietat comuna dintre d'una mateixa classe de comunicació i per tant es pot parlar tant de classes recurrents com transitòries.

Definició 3.4.5. Sigui $i, j \in I$. Es defineixen f_{ii}^* i f_{ij}^* com:

$$f_{ii}^* = \sum_{n=1}^{\infty} f_{ii}^n, \qquad f_{ij}^* = \sum_{n=1}^{\infty} f_{ij}^n$$

Encara hi ha una altra manera de caracteritzar el concepte de recurrència.

Sigui Q_{ii} la probabilitat de començar a l'estat i i retornar-hi infinitament sovint. Així doncs, l'estat i serà **recurrent** o **transitori** si $Q_{ii} = 1$ o $Q_{ii} = 0$ respectivament. Cal notar que Q_{ii} només pot prendre el valor d'1 o de 0.

Equivalentment doncs, es defineix $Q_{ij} = f_{ij}^* Q_{jj}$, on s'entèn que Q_{ij} és la probabilitat de visitar l'estat j infinitament sovint començant a l'estat i.

Proposició 3.4.3. Si $i \leftrightarrow j$ i la classe és recurrent, aleshores $f^*_{ij} = 1$

Proposició 3.4.4. Si $i \leftrightarrow j$ i la classe és recurrent, aleshores $Q_{ij} = 1$.

Demostració 3.4.4. Es compleix que $Q_{ij} = f^*_{ij} Q_{jj}$. Com que j és un estat recurrent es té que $Q_{jj} = 1$. Per altra banda, per la proposició anterior també es té que $f^*_{ij} = 1$, i per tant $Q_{ij} = 1$. $\square$

Entendre la recurrència d'aquesta manera serà molt útil per demostrar el teorema següent.

Teorema 3.4.3. Tota classe recurrent és tancada.

Demostració 3.4.5. Sigui C una classe no tancada. Per tant, $\exists\ i \in C, j \notin C$ i $m \geq 1$ tal que $P^m_{ij} > 0$.

Així doncs, hi ha una probabilitat positiva que el procés arribi a l'estat j. Si això passa, el procés no podrà tornar a l'estat i, ja que suposem que $i \to j$ i que $j \notin C$ però si hi tornés es tindria que $i \leftrightarrow j$ i per tant que $j \in C$ i s'arribaria a una contradicció.

El fet que un cop el procés arribi a l'estat j ja no podrà retornar a l'estat i indica que $Q_{ii} < 1 \Rightarrow Q_{ii} = 0$, ja que Q_{ii} només pot valer 0 o 1.

Per tant, i no és un estat recurrent i es conclou que C no serà una classe recurrent. $\square$

Teorema 3.4.4. Tota classe de comunicació tancada i finita és recurrent.

Demostració 3.4.6. Suposem que C és tancat i finit i que la cadena $\{X_n\}$ comença a C. Aleshores per algun estat $i \in C$ es té:

$$0 < \mathbb{P}\{X_n = i, \quad \text{per inifites } n\} = \sum_n \mathbb{P}\{X_n = i, X_m \neq i\ \forall m < n\} \cdot Q_{ii}$$

D'aquesta desigualtat s'extreu que $Q_{ii} > 0 \Rightarrow Q_{ii} = 1$. Per tant l'estat i no pot ser transitori. Com un estat només pot ser o bé transitori o bé recurrent, i serà recurrent i com que $i \in C$ i la recurrència és una propietat comuna al llarg de la mateixa classe, C també serà recurrent. $\square$.

3.4.2 Criteris de Recurrència

En aquesta secció tractarem dos teoremes que resulten especialment útils per determinar si una cadena de Markov és transitòria o recurrent i els veurem aplicats en un exemple.

Abans de res, introduïm unes definicions que farem servir per demostrar els teoremes següents.

Definició 3.4.6. Sigui $\{X_n\}$ un procés de Markov, C una classe recurrent i i un estat transitori. Aleshores $\boldsymbol{\pi_i(C)}$ és la probabilitat que començant a l'estat i, a la llarga el procés acabi sent absorbit per la classe C.

Definició 3.4.7. Es té que $\boldsymbol{\pi_i^n(C)}$ és la probabilitat que el procés entri i sigui absorbit per la classe recurrent C per primer cop a la n-èssima transició començant des de l'estat transitori i.

Notació 1. S'entèn $T \subseteq I$ com el conjunt d'estats transitoris.

Propietats 3.4.1. Veiem algunes propietats d'aquestes definicions.

1. $\pi_i(C) = \displaystyle\sum_{n=1}^{\infty} \pi_i^n(C) \leq 1$
2. $\pi_i^1(C) = \displaystyle\sum_{j \in C} P_{ij}$
3. $\pi_i^n(C) = \displaystyle\sum_{j \in T} P_{ij}\pi_j^{n-1}(C), \quad n \geq 2$

La propietat 1 es pot reescriure amb ajuda de les altres dues:

$$\pi_i(C) = \pi_i^1(C) + \sum_{n=2}^{\infty} \pi_i^n(C) = \pi_i^1(C) + \sum_{n=2}^{\infty} \sum_{j \in T} P_{ij}\pi_j^{n-1}(C) = \pi_i^1(C) + \sum_{j \in T} P_{ij} \sum_{n=2}^{\infty} \pi_j^{n-1}(C) \Rightarrow$$

$$\pi_i(C) = \pi_i^1(C) + \sum_{j \in T} P_{ij}\pi_j(C), \qquad i \in T$$

Teorema 3.4.5. Sigui $\{X_n\}$ una cadena de Markov $(\lambda, \mathbf{P})$ irreductible i $I \subseteq \mathbb{N}$ el seu conjunt d'estats. Una condició necessària i suficient per tal que $\{X_n\}$ sigui transitori és que el sistema d'equacions

$$\sum_{j \in I} P_{ij} y_j = y_i, \quad i \neq 0$$

tingui una solució no constant acotada.

Demostració 3.4.7. Sigui la matriu de transició del procés de la següent manera,

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ P_{20} & P_{21} & P_{22} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

i l'associem amb la nova matriu de transició

$$\tilde{\mathbf{P}} = \begin{bmatrix} 1 & 0 & 0 & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ P_{20} & P_{21} & P_{22} & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}$$

En aquesta nova matriu l'estat 0 s'ha convertit en una barrera absorbent mentre que la resta de probabilitats romanen inalterades.

Procedirem la demostració suposant que el procés és transitori i mostrant una solució no constant acotada. Al ser $\mathbf{P}$ transitori, es compleix que $f_{j0}^* < 1$ per alguna $j \neq 0$, ja que sinó l'estat 0 seria recurrent i arribaríem en una contradicció.

Com $\tilde{\mathbf{P}}$ té com a estat 0 una barrera absorbent, es complirà que $\tilde{\pi}_0(C_0) = 1$. Pel mateix argument d'abans, també se satisfarà que $\tilde{\pi}_j(C_0) = f_{j0}^* < 1$ per alguna $j \neq 0$. Per últim doncs es tindrà que:

$$\tilde{\pi}_i(C_0) = \sum_{j \in I} \tilde{P}_{ij} \tilde{\pi}_j(C_0), \quad \forall i$$

Per tant,

$$\tilde{\pi}_i(C_0) = \sum_{j \in I} P_{ij} \tilde{\pi}_j(C_0), \quad \forall i \neq 0$$

I per tant la solució acotada desitjada no constant, donat que $\tilde{\pi}_0(C_0) = 1$ i $\tilde{\pi}_j(C_0) < 1$, serà

$$y_j = \tilde{\pi}_j(C_0), \quad j = 0, 1, 2, \cdot$$

Suposem ara que tenim una solució acotada $\{y_i\}$ per l'equació del teorema.

$$\sum_{j \in I} \tilde{P}_{ij} y_j = y_i, \quad \forall i \geq 0$$

Iterant, es té que $\forall i \geq 0$, $\forall n \geq 1$,

$$\sum_{j \in I} \tilde{P}_{ij}^n y_j = y_i$$

Si la cadena és recurrent, aleshores

1) $\lim_{n \to \infty} \tilde{P}_{i0}^n = 1$

2) $\sum_{j \neq 0} \tilde{P}_{ij}^n y_j \leq M(1 - \tilde{P}_{i0}^n) \to 0 \quad$ quan $n \to \infty$

Notem que 1) es compleix perquè $\tilde{P}^n_{i0} = \sum_{k=1}^{n} f^k_{i0}$. Al fer el límit, es té que $\lim_{n\to\infty} \tilde{P}^n_{i0} = \sum_{k=1}^{\infty} f^k_{i0} = f^*_{i0} = 1$ ja que suposem que la cadena és recurrent.

Notem que a 2) M és una cota de $\{y_i\}$.

Per tant, reescrivint y_i es veu que

$$y_i = \sum_{j\neq 0} \tilde{P}^n_{ij} y_j + \tilde{P}^n_{i0} y_0 \to y_0 \quad \text{quan } n \to \infty$$

Així doncs, $y_i = y_0$ $\forall i$ i $\{y_i\}$ és constant. $\square$

Teorema 3.4.6. Una condició suficient per a què una cadena de Markov irreductible sigui recurrent és que existeixi una seqüència $\{y_i\}_{i\in I}$ amb $y_i \to \infty$ tal que

$$\sum_{j\in I} P_{ij} y_j \leq y_i \quad \text{per } i \neq 0$$

Demostració 3.4.8. Al llarg de la demostració entendrem la matriu $\tilde{\mathbf{P}}$ igual que s'ha definit a la demostració 3.4.7.

Per tant, es compleix que

$$\sum_{j\in I} \tilde{P}_{ij} y_j \leq y_i, \quad \forall i$$

Iterant aquesta desigualtat, s'arriba a que

$$\sum_{j\in I} \tilde{P}^m_{ij} y_j \leq y_i$$

Cal notar que com $\forall b$, $z_i = y_i + b$ també satisfà la desigualtat del teorema, juntament amb el fet que $\lim y_i = \infty$, podem assumir que $y_i > 0$ $\forall i \geq 0$.

Donat un $\epsilon > 0$ escollim $M(\epsilon)$ tal que $\frac{1}{y_i} \leq \epsilon$ per $i \geq M(\epsilon)$. Reescrivint la darrera desigualtat es té que

$$\sum_{j=0}^{M-1} \tilde{P}^m_{ij} y_j + \sum_{j\geq M} \tilde{P}^m_{ij} y_j \leq y_i \iff \sum_{j=0}^{M-1} \tilde{P}^m_{ij} y_j + \min_{r\geq M}\{y_r\} \sum_{j\geq M} \tilde{P}^m_{ij} \leq y_i \iff$$

$$\iff \sum_{j=0}^{M-1} \tilde{P}^m_{ij} y_j + \min_{r\geq M}\{y_r\} \left(\sum_{j\in I} \tilde{P}^m_{ij} - \sum_{j=0}^{M-1} \tilde{P}^m_{ij} \right) \leq y_i \overset{a}{\iff}$$

$$\overset{a}{\iff} \sum_{j=0}^{M-1} \tilde{P}^m_{ij} y_j + \min_{r\geq M}\{y_r\} \left(1 - \sum_{j=0}^{M-1} \tilde{P}^m_{ij} \right) \leq y_i$$

Notem que a la doble implicació a s'ha utilitzat que per un i fix, se satisfà $\sum_{j\in I} \tilde{P}_{ij}^m = 1$. També que a la demostració 3.4.7. s'ha vist que $\lim_{n\to\infty} \tilde{P}_{ij}^n = 0$, per $j \neq 0$.

Així doncs, fent el límit per $m \to \infty$ a la darrera desigualtat, per a cada i fixe s'obté:

$$\lim_{m\to\infty} \tilde{P}_{i0}^m y_0 + \min_{r\geq M}\{y_r\} \left(1 - \lim_{m\to\infty} \tilde{P}_{i0}^m\right) \leq y_i$$

Com l'estat 0 és una barrera absorbent, $\lim_{m\to\infty} \tilde{P}_{i0}^m = \tilde{\pi}_i(C_0)$, i aquesta desigualtat es pot reescriure com:

$$\tilde{\pi}_i(C_0)y_0 + \min_{r\geq M}\{y_r\} \left(1 - \tilde{\pi}_i(C_0)\right) \leq y_i \iff 1 - \tilde{\pi}_i(C_0) \leq \frac{1}{\min_{r\geq M}\{y_r\}}(y_i - \tilde{\pi}_i(C_0)y_0) \leq \epsilon(y_i - \tilde{\pi}_i(C_0)y_0)$$

Com ϵ és arbitrari i $\tilde{\pi}_i(C_0) \leq 1$ es té que $\tilde{\pi}_i(C_0) = 1 \ \forall i$, el que demostra que el procés és efectivament recurrent. □

Exemple 3.4.1. ***Procés discret al fer una cua***

Tornem a rependre aquest exemple per tal d'il·lustrar l'aplicabilitat dels teoremes anteriors. Treballarem sota les hipòtesis que $0 < a_0 < 1$ i que $a_0 + a_1 < 1$. Així doncs, recordant el que s'ha explicat a l'exemple 3.3.2. amb aquestes hipòtesis ens assegurem que el procés sigui irreductible, condició necessària per tal de poder aplicar els dos teoremes anteriors.

Basant-nos en ells, veurem que si el valor esperat de nous clients, $\sum_{k=0}^{\infty} k a_k$, és major que 1, aleshores a mesura que passi el temps la cua augmentarà sense límit, és a dir, la cadena de Markov serà transitòria. En el cas contrari, quan el valor esperat de nous clients no sigui major que 1, la cua arribarà a un estat d'equilibri, és a dir, la cadena de Markov serà recurrent.

Recordem que

$$\mathbf{P} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & \dots \\ a_0 & a_1 & a_2 & a_3 & \dots \\ 0 & a_0 & a_1 & a_2 & \dots \\ 0 & 0 & a_0 & a_1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

$$\text{on } a_k > 0 \text{ i } \sum_{k=0}^{\infty} a_k = 1$$

A) Si $\sum_{k=0}^{\infty} k a_k > 1$, volem veure que el sistema d'equacions $\sum_{j=0}^{\infty} P_{ij} y_j = y_i$, on $i \neq 0$, admet una solució acotada no constant i pel teorema 3.4.5. el procés serà transitori. Si prenem y_j de forma exponencial, és a dir, $y_j = \xi^j$, aleshores,

$$\sum_{j=0}^{\infty} P_{ij}\xi^j = \sum_{j=i-1}^{\infty} a_{j-i+1}\xi^j = \xi^i$$

Notem que en aquest procés, quan es vol passar a un estat de valor més petit, aquest només pot ser inferior en una sola unitat.

Treballant amb la darrera expressió s'arriba a que,

$$\sum_{j=i-1}^{\infty} a_{j-i+1}\xi^{j-i} = 1 \iff \sum_{j=i-1}^{\infty} a_{j-i+1}\xi^{j-i+1} = \xi, \quad \text{on } i \neq 0$$

i es defineix la funció $f(\xi) = \sum_{k=0}^{\infty} a_k \xi^k$.

D'aquestes darreres expressions s'extreu que:

(a) $f(0) = a_0 > 0$

(b) $f(1) = \sum_{k=0}^{\infty} a_k = 1$

(c) $f'(1) = \sum_{k=0}^{\infty} k a_k > 1$

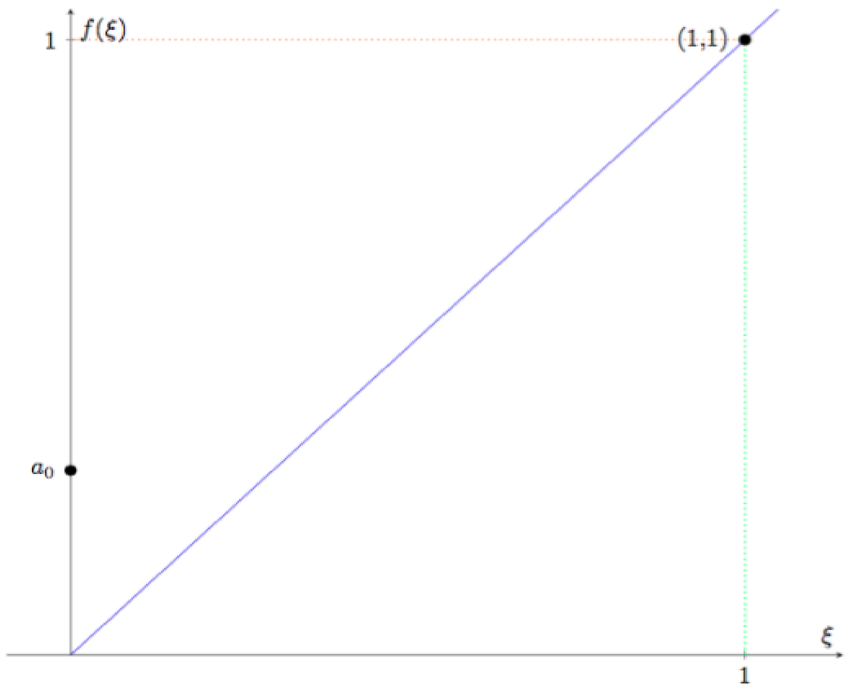

Figura 3: *Representació gràfica d'algunes funcions per a entendre millor la demostració.*

Per tant, es té que $\exists\ \xi_0, 0 < \xi_0 < 1 \mid f(\xi_0) = \xi_0$. Aquest fet es visualitza amb més facilitat a la Figura 3. Veiem que per anar del punt a_0 al $(1,1)$, la funció arribarà a aquest darrer punt passant entre les rectes blava, $y = \xi$, i verda, $\xi = 1$, per tal que tingui un pendent que satisfaci que $f'(1) > 1$. Per tant, per unir aquests dos punts, en algun moment s'haurà de tallar la recta $y = \xi$ i per tant podem assegurar que $\exists\ \xi_0, 0 < \xi_0 < 1 \mid f(\xi_0) = \xi_0$.

Així doncs, es té que el vector $y_j = \xi_0^j$ per $i = 0, 1, \cdots$ és la solució acotada no constant desitjada i mitjançant el teorema 3.4.5. queda demostrat que en aquest cas el procés és transitori.

B) Si $\sum_{k=0}^{\infty} k a_k \leq 1$, procedim a aplicar el teorema 3.4.6. prenent $y_j = j$. Tenint en compte que $i \neq 0$, aleshores se satisfà el següent:

$$\sum_{j=0}^{\infty} P_{ij} j = \sum_{j=i-1}^{\infty} a_{j-i+1} j \overset{b}{=} \sum_{j=i-1}^{\infty} a_{j-i+1}(j-i+1) + i - 1 = \sum_{k=0}^{\infty} k a_k - 1 + i \overset{c}{\leq} i$$

Cal esmentar però què ha passat en les igualtats anteriors. En la igualtat b veiem que:

$$\sum_{j=i-1}^{\infty} a_{j-i+1} j = \sum_{j=i-1}^{\infty} a_{j-i+1}(j-i+1+i-1) = \sum_{j=i-1}^{\infty} a_{j-i+1}(j-i+1) + \sum_{j=i-1}^{\infty} a_{j-i+1}(i-1) =$$

$$= \sum_{j=i-1}^{\infty} a_{j-i+1}(j-i+1) + (i-1) \sum_{j=i-1}^{\infty} a_{j-i+1} = \sum_{j=i-1}^{\infty} a_{j-i+1}(j-i+1) + (i-1)$$

Per altra banda, es veu que la desigualtat c és certa ja que:

$$\sum_{k=0}^{\infty} k a_k \leq 1 \Rightarrow \sum_{k=0}^{\infty} k a_k - 1 \leq 0 \Rightarrow \sum_{k=0}^{\infty} k a_k - 1 + i \leq i$$

Així doncs queda vist que pel cas $\sum_{k=0}^{\infty} k a_k \leq 1$ el teorema 3.4.6. es compleix i per tant el procés és recurrent. □

3.5 Probabilitats d'Absorció

Definició 3.5.1. Sigui $\{X_n\}$ una cadena de Markov $(\lambda, \mathbf{P})$. El **temps d'arribada** a un subconjunt $A \subseteq I$ és la variable aleatòria $H^A : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ definida:

$$H^A(\omega) = \inf\{n \geq 0 : X_n(\omega) \in A\}$$

Observació 3.5.1. Recordem que l'ínfim del conjunt buit $\emptyset$ és ∞.

Definició 3.5.2. Sigui h_i^A la probabilitat que la cadena $\{X_n\}$ arribi al conjunt A començant per l'estat i:

$$h_i^A = \mathbb{P}\{H^A < \infty | X_0 = i\}$$

Definició 3.5.3. En el cas que A sigui una classe tancada, h_i^A s'anomena la **probabilitat d'absorció**.

Definició 3.5.4. El temps mitjà que triga la cadena $\{X_n\}$ per arribar al conjunt A, k_i^A, ve donat per:

$$k_i^A = \mathbb{E}\left[H^A | X_0 = i\right] = \sum_{n<\infty} n\mathbb{P}\{H^A = n\} + \infty\mathbb{P}\{H^A = \infty\}$$

Observació 3.5.2. Per convenció, a la definició anterior s'ha pres que $\infty \cdot 0 = 0$.

Notació 2. D'una manera més lleugera, escriurem

1. $h_i^A = \mathbb{P}\{\text{hit } A | X_0 = i\} = \mathbb{P}_i(\text{hit } A)$
2. $k_i^A = \mathbb{E}\left[\text{temps per fer hit a } A | X_0 = i\right] = \mathbb{E}_i(\text{temps per fer hit a } A)$

A continuació donarem forma a aquests nous conceptes mitjançant un exemple.

Exemple 3.5.1. Sigui $\{X_n\}$ una cadena de Markov (λ, $\mathbf{P}$) amb $I = \{0, 1, 2, 3, 4\}$ i on

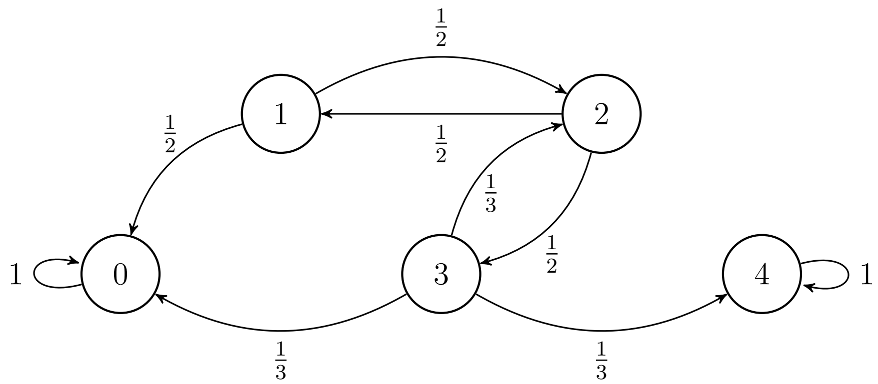

Figura 4: *Representació del comportament de la cadena de Markov descrita per la matriu* $\boldsymbol{P}$.

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

Fent una primera anàlisi de **P** veiem que tant $\{0\}$ com $\{4\}$ són classes tancades, així que ambdós estats 1 i 4 són absorbents. Així doncs, ens podem plantejar les següents preguntes: començant a l'estat 2, quina és la probabilitat d'absorció a 4? Quant trigaria la cadena per a ser absorbida pels estats 0 o 4? Calculem-ho:

$$h_2 = \mathbb{P}_2(\text{hit } 4)\ , \quad k_2 = \mathbb{E}_2(\text{temps per fet hit a } \{0,4\})$$

Al tractar-se tant 0 com 4 d'estats absorbents i al voler calcular h_2 per arribar a 4, es té que $h_0 = 0$ mentre que $h_4 = 1$. Estudiant **P** i tenint en compte que treballem amb una cadena de Markov i que a cada nou instant només es té en consideració l'estat justament anterior, s'arriba a

$$\left\{\begin{array}{c} h_1 = \frac{1}{2}h_0 + \frac{1}{2}h_2 \\ h_2 = \frac{1}{2}h_1 + \frac{1}{2}h_3 \\ h_3 = \frac{1}{3}h_0 + \frac{1}{3}h_2 + \frac{1}{3}h_4 \end{array}\right. \overset{h_0=0,\ h_4=1}{=} \left\{\begin{array}{c} h_1 = \frac{1}{2}h_2 \\ h_2 = \frac{1}{2}h_1 + \frac{1}{2}h_3 \\ h_3 = \frac{1}{3}h_2 + \frac{1}{3} \end{array}\right. = \left\{\begin{array}{c} h_2 = \frac{1}{4}h_2 + \frac{1}{2}h_3 \\ h_3 = \frac{1}{3}h_2 + \frac{1}{3} \end{array}\right.$$

$$\Rightarrow h_2 = \frac{1}{4}h_2 + \frac{1}{2}\left(\frac{1}{3}h_2 + \frac{1}{3}\right) \iff h_2 = \frac{2}{7}$$

Pel mateix raonament que 0 i 4 són estats absorbents, al calcular k_2 se satisfarà que $k_0 = k_4 = 0$. Tornant a extreure les dades de **P** s'arriba a:

$$\left\{\begin{array}{c} k_1 = 1 + \frac{1}{2}k_0 + \frac{1}{2}k_2 \\ k_2 = 1 + \frac{1}{2}k_1 + \frac{1}{2}k_3 \\ k_3 = 1 + \frac{1}{3}k_0 + \frac{1}{3}k_2 + \frac{1}{3}k_4 \end{array}\right. \overset{k_0=k_4=0}{=} \left\{\begin{array}{c} k_1 = 1 + \frac{1}{2}k_2 \\ k_2 = 1 + \frac{1}{2}k_1 + \frac{1}{2}k_3 \\ k_3 = 1 + \frac{1}{3}k_2 \end{array}\right. = \left\{\begin{array}{c} k_2 = 1 + \frac{1}{2}\left(1 + \frac{1}{2}k_2\right) + \frac{1}{2}k_3 \\ k_3 = 1 + \frac{1}{3}k_2 \end{array}\right.$$

$$\Rightarrow k_2 = 1 + \frac{1}{2}\left(1 + \frac{1}{2}k_2\right) + \frac{1}{2}\left(1 + \frac{1}{3}k_2\right) \iff k_2 = \frac{24}{7}$$

Cal notar que pel càlcul de k_2 s'ha afegit un 1 que representa el primer pas.

Aquest exemple mostra el funcionament conceptual d'aquestes eines. No obstant, les quantitats de h_i^A i k_i^A es poden calcular explícitament d'una manera genèrica mitjançant equacions lineals associades amb els elements de **P**, com es pot apreciar en l'exemple d'una manera indirecta.

Teorema 3.5.1. El vector de les probabilitats d'arribada, $h^A = \left(h_i^A : i \in I\right)$ és la solució no negativa minimal del sistema d'equacions lineals:

$$\left\{\begin{array}{ll} h_i^A = 1 & \text{si } i \in A \\ h_i^A = \displaystyle\sum_{j\in I} P_{ij}h_j^A & \text{si } i \notin A \end{array}\right.$$

Observació 3.5.3. Que h^A sigui la solució minimal significa que si $x = (x_i : i \in I)$ també és una solució amb $x_i \geq 0\ \forall i$, aleshores $x_i \geq h_i^A\ \forall i$.

Demostració 3.5.1. Primer procedim a demostrar que h^A satisfà el sistema d'equacions.

i) Si $X_0 = i \in A$, aleshores

$$H^A = 0, \text{i per tant } h_i^A = 1$$

ii) Si $X_0 = i \notin A$, aleshores $H^A \geq 1$ i per la propietat de Markov es complirà que

$$\mathbb{P}\{H^A < \infty | X_1 = j, X_0 = i\} = \mathbb{P}\{H^A < \infty | X_0 = j\} = h_j^A$$

Per altra banda també es compleix que

$$h_i^A = \mathbb{P}\{H^A < \infty \ | X_0 = i\} = \sum_{j \in I} \mathbb{P}\{H^A < \infty, X_1 = j | X_0 = i\} \stackrel{\text{a}}{=}$$

$$\stackrel{\text{a}}{=} \sum_{j \in I} \mathbb{P}\{H^A < \infty | X_1 = j, X_0 = i\} \mathbb{P}\{X_1 = j | X_0 = i\} = \sum_{j \in I} P_{ij} h_j^A$$

A la igualtat a s'ha fet servir la definició de probabilitat condicionada.

A continuació procedim a demostrar que efectivament h^A és la solució minimal. Suposem que $x = (x_i : i \in I)$ és una solució de les equacions lineals del teorema.

i) Si $i \in A$, aleshores $h_i^A = x_i = 1$.

ii) Si $i \notin A$, aleshores

$$x_i = \sum_{j \in I} P_{ij} x_j = \sum_{j \in A} P_{ij} + \sum_{j \notin A} P_{ij} x_j$$

Substituint per x_j s'obté que:

$$x_i = \sum_{j \in A} P_{ij} + \sum_{j \notin A} P_{ij} \left(\sum_{k \in A} P_{jk} + \sum_{k \notin A} P_{jk} x_k \right) =$$

$$= \mathbb{P}\{X_1 \in A | X_0 = i\} + \mathbb{P}\{X_1 \notin A, X_2 \in A | X_0 = i\} + \sum_{j \notin A} \sum_{k \notin A} P_{ij} P_{jk} x_k$$

Iterant les substitucions de x, després de n passos s'arriba a:

$$x_i = \mathbb{P}\{X_1 \in A | X_0 = i\} + \ldots + \mathbb{P}\{X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A | X_0 = i\} +$$
$$+ \sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} P_{ij_1} P_{j_1 j_2} \ldots P_{j_{n-1} j_n} x_{j_n} = \mathbb{P}\{H^A \leq n | X_0 = i\} + \sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} P_{ij_1} P_{j_1 j_2} \ldots P_{j_{n-1} j_n} x_{j_n}$$

Si x és no negatiu, x_{j_n} tampoc ho serà. Així doncs, la suma de tots els elements de la dreta de la igualtat serà igual a $\mathbb{P}\{H^A \leq n|X_0 = i\}$. Per tant, es complirà que $x_i \geq \mathbb{P}\{H^A \leq n|X_0 = i\}\ \forall\ n$ i per tant

$$x_i \geq \lim_{n\to\infty} \mathbb{P}\{H^A \leq n|X_0 = i\} = \mathbb{P}\{H^A < \infty|X_0 = i\} = h_i \quad \square$$

Teorema 3.5.2. El vector de temps esperat d'arribada $k^A = (k_i^A : i \in I)$ és la solució no negativa minimal del sistema d'equacions lineals

$$\begin{cases} k_i^A = 0 & \text{si } i \in A \\ k_i^A = 1 + \displaystyle\sum_{j\notin A} P_{ij}k_j^A & \text{si } i \notin A \end{cases}$$

Demostració 3.5.2. Primer procedim a demostrar que k^A satisfà el sistema d'equacions.

· Si $X_0 = i \in A$, aleshores

$$H^A = 0, \text{i per tant } k_i^A = 0$$

· Si $X_0 = i \notin A$, aleshores $H^A \geq 1$ i per la propietat de Markov es complirà que

$$\mathbb{E}\left[H^A|X_1 = j, X_0 = i\right] = \mathbb{E}\left[H^A - 1|X_1 = j\right] + 1 = \mathbb{E}\left[H^A|X_0 = j\right] + 1$$

Notem que com estem mesurant el temps se suma l'1 ja que el procés ja ha fet un pas.

Per altra banda també es compleix que

$$k_i^A = \mathbb{E}\left[H^A|X_0 = i\right] \stackrel{\text{a}}{=} \sum_{j\in I} \mathbb{E}\left[H^A|X_1 = j, X_0 = i\right] \mathbb{P}\{X_1 = j|X_0 = i\} =$$

$$= \sum_{j\in I} \left(1 + \mathbb{E}\left[H^A|X_0 = j\right]\right) \mathbb{P}\{X_1 = j|X_0 = i\} = \sum_{j\in I} P_{ij} + \sum_{j\in I} P_{ij}k_j^A \stackrel{\text{b}}{=}$$

$$\stackrel{\text{b}}{=} 1 + \sum_{j\notin A} P_{ij}k_j^A$$

En síntesi, a la igualtat a s'ha utilitzat que:

$$\mathbb{E}\left[H^A\right] = \sum_{n=0}^{\infty} n \cdot \mathbb{P}\{H^A = n\} = \sum_{n=0}^{\infty} n \cdot \left(\sum_{j\in I} \mathbb{P}\{H^A = n|X_1 = j\}\mathbb{P}\{X_1 = j\}\right) =$$

$$= \sum_{j\in I} \mathbb{P}\{X_1 = j\} \sum_{n=0}^{\infty} n \cdot \mathbb{P}\{H^A|X_1 = j\} = \sum_{j\in I} \mathbb{P}\{X_1 = j\}\mathbb{E}\left[H^A|X_1 = j\right]$$

A la igualtat b s'ha fet servir que:

la proposició 3.1.1. que diu que $\displaystyle\sum_{j\in I} P_{ij} = 1$ i també a que $k_i^A = 0\ \forall i \in A$.

A continuació procedim a demostrar que efectivament k^A és la solució minimal.
Sigui $y = (y_i : i \in I)$ una altra solució de les equacions lineal del teorema.
· Si $i \in A$, aleshores $k_i^A = y_i = 0$
· Si $i \notin A$, aleshores

$$y_i = 1 + \sum_{j \notin A} P_{ij} y_j$$

Substituint per y_j s'obté que:

$$y_i = 1 + \sum_{j \notin A} P_{ij} \left(1 + \sum_{k \notin A} P_{jk} y_k \right) = \mathbb{P}\{H^A \geq 1 | X_0 = i\} + \mathbb{P}\{H^A \geq 2 | X_0 = i\} + \sum_{j \notin A} \sum_{k \notin A} P_{ij} P_{jk} y_k$$

Iterant les substitucions de y, després d'n passos s'arriba a:

$$y_i = \mathbb{P}\{H^A \geq 1 | X_0 = i\} + \ldots + \mathbb{P}\{H^A \geq n | X_0 = i\} + \sum_{j_1 \notin A} \ldots \sum_{j_n \notin A} P_{ij_1} P_{j_1 j_2} \ldots P_{j_{n-1} j_n} y_{j_n}$$

Suposant que la solució y es no negativa, aleshores,

$$y_i \geq \mathbb{P}\{H^A \geq 1 | X_0 = i\} + \ldots + \mathbb{P}\{H^A \geq n | X_0 = i\}$$

I fent el límit quan $n \to \infty$,

$$y_i \geq \sum_{n=1}^{\infty} \mathbb{P}\{H^A \geq n | X_0 = i\} = \mathbb{E}\{H^A | X_0 = i\} = k_i^A \quad \square$$

4 Cadenes de Markov d'Ordre Superior

4.1 Introducció

Fins a aquest punt del treball s'han estudiat les cadenes de Markov, que per definició hem vist que són processos "sense memòria" que només tenen en compte la situació del procés en l'estat justament anterior. No obstant, el concepte de cadenes de Markov d'ordre superior el que fa és no només fixar-se en l'estat anterior sinó que mira més enllà en el passat.

Definició 4.1.1. Sigui $\{X_n\}$ un procés. Direm que és una **cadena de Markov d'ordre** k si té en compte els k instants anteriors del procés.

Observació 4.1.1. Les cadenes de Markov estudiades fins ara serien d'ordre 1.

Notació 3. En aquesta secció s'utilitzarà una notació una miqueta modificada a la utilitzada fins ara per tal d'agilitzar l'escriptura per aquest tipus de cadenes de Markov.

Suposem que tenim una seqüència de dades categòriques on cada punt, $X^{(n)}$, pot prendre valors del conjunt $M = \{1, 2, \ldots, m\}$. Cal notar que m és finit i per tant la seqüència té m estats possibles. En el model més convencional de cadena de Markov d'ordre k, que és el model on es fa el paral·lisme més directe a les cadenes de Markov, es tenen $(m-1)m^k$ paràmetres. Clarament, el número de paràmetres augmenta exponencialment amb l'ordre de la cadena, el que complica moltíssim els càlculs. Per aquesta raó aquest model no és gaire popular.

Observació 4.1.2. En una cadena de Markov d'ordre 1 cadascuna de les probabilitats de transició entre dos estats serà un paràmetre del model. La dimensió de la matriu de transició serà de $m \times m$. Tot i que una matriu de transició tindrà m^2 elements, com hi ha la condició que la suma d'element de cada fila ha de donar 1 tindrà $(m-1)m$ paràmetres. Extenent això per quan la cadena de Markov és d'ordre k, totes les probabilitats es podran descriure per un vector d'ordre $m^k \times m$. Això es deu a que si mirem els k instants anteriors a l'estat i d'un procés amb un conjunt de valors d'ordre m, hi haurà m^k combinacions possibles dels k instants anteriors. En canvi, l'instant i només podrà assolir m valors diferents. Per això la dimensió del vector és de $m^k \times m$. De nou, per l'instant i només hi haurà $m-1$ graus de llibertat ja que existeix la condició que la suma de probabilitats donarà 1. Per tant, el nombre de paràmetres llavors és efectivament de $(m-1)m^k$.

4.2 Característiques i Propietats del Model

4.2.1 El Nou Model

Aquest nou model amb el que treballarem va ser proposat per Raftery i té com a objectiu mantenir un número baix de paràmetres.

El model es basa en afegir un paràmetre extra per a cada estat enrere que vulguem mirar i s'escriu de la manera següent:

$$\mathbb{P}\{X^{(n)} = j_0 | X^{(n-1)} = j_1, \cdots, X^{(n-k)} = j_k\} = \sum_{i=1}^{k} \lambda_i q_{j_0 j_i}$$

$$\text{on } \sum_{i=1}^{k} \lambda_i = 1$$

Entenem els λ_i com a pesos i assumim que són no negatius i que se satisfà $0 \leq q_{j_0 j_i} \leq 1$ al tractar-se de probabiltiats.

Si prenem Q_i com a matrius de transició, el model es pot reescriure de la manera següent:

$$\mathbf{X}^{(n+k+1)} = \sum_{i=1}^{k} \lambda_i Q_i \mathbf{X}^{(n+k+1-i)}$$

$\mathbf{X}^{(n+k+1-i)}$ s'entèn com la distribució de probabilitat dels estats en l'instant $(n + k + 1 - i)$.

Cal notar que com $\sum_{i=1}^{k} \lambda_i = 1$ i Q_i són matrius de transició, cada element del vector $\mathbf{X}^{(n+k+1)}$ tindrà un valor entre 0 i 1 i que la suma de cada element del vector serà de 1.

Observació 4.2.1. També s'ha de mencionar que en aquest nou model, el nombre de paràmetres serà de $k + km(m - 1)$. Veiem per què. Abans hem dit que aquest model funciona afegint un paràmetre extra per a cada instant anterior al que ens remuntem. Al ser una cadena de Markov d'ordre k, ens remuntem a k instants anteriors, el que afegirà k paràmetres. Per altra banda, cada matriu de transició tindrà m^2 elements, però com hi ha la condició que la suma d'elements de cada fila ha de donar 1, hi haurà $m(m - 1)$ paràmetres. Comptabilitzant-ho tot plegat per k matrius de transició, s'arriba al valor final de $k + km(m - 1)$ paràmetres.

Veiem que tot sovint es dóna el cas en que $Q \equiv Q_1 = Q_2 = \ldots = Q_k$. Quan això passa, el model queda de la següent forma:

$$\mathbf{X}^{(n+k+1)} = \sum_{i=1}^{k} \lambda_i Q \mathbf{X}^{(n+k+1-i)}$$

on ara només tindrà $k + m(m - 1)$ paràmetres.

Havent presentat el model, cal recalcar que es treballa suposant que $\mathbf{X}^{(n+k+1)}$ depèn de $\mathbf{X}^{(n+i)}$, on $i = 1, 2, \ldots, k$, és a dir, de tots els k instants anteriors i no només de l'instant anterior k. Aquesta dependència es transmet a través de la matriu de transició Q_i i del pes λ_i. Recordem que cada Q_i és una matriu estocàstica no negativa on els elements de les seves columnes sumen 1.

4.2.2 Propietats

Abans d'explicar el mètode per estimar els paràmetres del model, presentem unes proposicions necessàries.

Proposició 4.2.1. Sigui Q_i irreductible i $\lambda_i > 0$ tal que

$$0 \leq \lambda_i \leq 1 \text{ i } \sum_{i=1}^{k} \lambda_i = 1$$

Aleshores el model proposat té una distribució estacionària $\bar{\mathbf{X}}$ quan $n \to \infty$ independent dels vector d'estat inicial $\mathbf{X}^{(0)}, \mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(k-1)}$. La distribució estacionària $\bar{\mathbf{X}}$ també és l'única solució del sistema lineal d'equacions:

$$\left(I - \sum_{i=1}^{n} \lambda_i Q_i\right) \bar{\mathbf{X}} = \mathbf{0} \text{ i } \mathbf{1}^T \bar{\mathbf{X}} = 1$$

Aquest resultat es pot entendre d'una manera més laxa tal com ho expressa la següent proposició.

Proposició 4.2.2. Si Q_k és irreductible i aperiòdica, $\lambda_1, \lambda_k > 0$ i $\sum_{i=1}^{k} \lambda_i = 1$, aleshores el model té una distribució estacionària $\mathbf{X}$ que satisfà:

1. $\left(I - \sum_{i=1}^{n} \lambda_i Q_i\right) \bar{\mathbf{X}} = \mathbf{0} \text{ i } \mathbf{1}^T \bar{\mathbf{X}} = 1$
2. $\lim_{n \to \infty} \mathbf{X}^{(n)} = \mathbf{X}$

4.2.3 Estimació de Paràmetres

Procedim a presentar mètodes eficients per tal d'estimar els paràmetres del model: Q_i i λ_i per $i = 1, 2, \ldots, k$.

A l'hora d'estimar Q_i, és molt beneficiós interpretar Q_i com la matriu de transició del pas i-èssim de la seqüència categòrica de dades $\{X^{(n)}\}$. Treballant amb $\{X^{(n)}\}$ es poden comptabilitzar les freqüències de transició $\phi_{lj}^{(i)}$, que indiquen quants cops dintre de la seqüència es pot anar de l'estat l al j en i passes.

Definició 4.2.1. La matriu $\Phi^{(i)} = [\phi_{lj}^{(i)}]$

$$\Phi^{(i)} = \begin{bmatrix} \phi_{11}^{(i)} & \phi_{21}^{(i)} & \dots & \phi_{m1}^{(i)} \\ \phi_{12}^{(i)} & \phi_{22}^{(i)} & \dots & \phi_{m2}^{(i)} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_{1m}^{(i)} & \phi_{2m}^{(i)} & \dots & \phi_{mm}^{(i)} \end{bmatrix}$$

es defineix com la **matriu de freqüència de transició del pas i-èssim** de la seqüència categòrica de dades $\{X^{(n)}\}$.

Mitjançant la matriu $\Phi^{(i)}$ estimem $\hat{Q}_i = [\hat{q}_{lj}^{(i)}]$ amb

$$\hat{Q}_i = \begin{bmatrix} \hat{q}_{11}^{(i)} & \hat{q}_{21}^{(i)} & \dots & \hat{q}_{m1}^{(i)} \\ \hat{q}_{12}^{(i)} & \hat{q}_{22}^{(i)} & \dots & \hat{q}_{m2}^{(i)} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{q}_{1m}^{(i)} & \hat{q}_{2m}^{(i)} & \dots & \hat{q}_{mm}^{(i)} \end{bmatrix}$$

on

$$\hat{q}_{lj}^{(i)} = \begin{cases} \dfrac{\phi_{lj}^{(i)}}{\sum_{j=1}^{m} \phi_{lj}^{(i)}} & \text{si } \sum_{j=1}^{m} \phi_{lj}^{(i)} \neq 0 \\ 0 & \text{en cas contrari} \end{cases}$$

Observació 4.2.2. Cal notar que la complexitat computacional de la construcció de $\Phi^{(i)}$ augmenta proporcionalment de forma directa amb l'ordre k de la cadena de Markov que s'estigui tractant.

Abans de continuar, veiem que els estimadors proposats no són esbiaixats.

Proposició 4.2.3. Els estimadors que s'acaben de proposar satisfan que

$$\mathrm{E}\left[\phi_{lj}^{(i)}\right] = q_{lj}^{(i)} \mathrm{E}\left[\sum_{j=1}^{m} \phi_{lj}^{(i)}\right]$$

Demostració 4.2.1. Sigui $[q_{lj}^{(i)}]$ la matriu de transició del pas i-èssim, i P_l la probabilitat que el procés es trobi en l'estat l. Aleshores existeix una constant α tal que,

$$\mathrm{E}\left[\phi_{lj}^{(i)}\right] = \alpha \cdot P_l \cdot q_{lj}^{(i)}$$

i per tant

$$\mathrm{E}\left[\sum_{j=1}^{m} \phi_{lj}^{(i)}\right] = \alpha \cdot P_l \cdot \left(\sum_{j=1}^{m} q_{lj}^{(i)}\right) = \alpha \cdot P_l$$

Substituint, s'arriba a que

$$\mathrm{E}\left[\phi_{lj}^{(i)}\right] = q_{lj}^{(i)}\mathrm{E}\left[\sum_{j=1}^{m}\phi_{lj}^{(i)}\right] \square$$

Procedim a veure com estimar els paramètres λ_i.

Recordem que la proposició 4.2.1.dóna una condició suficient per a què la seqüència $\mathbf{X}^{(n)}$ convergeixi a una distribució estacionària $\mathbf{X}$. Si suposem que $\lim_{n\to\infty}\mathbf{X}^{(n)} = \bar{\mathbf{X}}$, aleshores $\bar{\mathbf{X}}$ es pot estimar calculant la proporció d'ocurrències de cada estat en la seqüència, que ho representarem per $\hat{\mathbf{X}}$. Si satisfés les equacions lineals de la proposició 4.2.1. s'esperaria que:

$$\sum_{i=1}^{k}\lambda_i\hat{Q}_i\hat{\mathbf{X}} \cong \hat{\mathbf{X}}$$

Per tant, una possible manera d'estimar els paràmetres $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_k)$ és considerar el següent problema de minimització:

$$\min_{\lambda}\left\{\left\|\sum_{i=1}^{k}\lambda_i\hat{Q}_i\hat{\mathbf{X}} - \hat{\mathbf{X}}\right\|\right\}$$

amb les condicions que

$$\sum_{i=1}^{k}\lambda_i = 1 \;\; \text{i} \;\; \lambda_i \geq 0, \forall i$$

on $||\cdot||$ representa una norma vectorial. En cas que es treballi amb $||\cdot||_\infty$ el problema de minimització es veurà modificat al següent:

$$\min_{\lambda}\left\{\max_{l}\left|\left[\sum_{i=1}^{k}\lambda_i\hat{Q}_i\hat{\mathbf{X}} - \hat{\mathbf{X}}\right]_l\right|\right\}$$

amb les condicions que

$$\sum_{i=1}^{k}\lambda_i = 1 \;\; \text{i} \;\; \lambda_i \geq 0, \forall i$$

Cal esmentar que $[\cdot]_l$ denota l'element l-èssim del vector. També que les condicions imposades a l'hora d'optimitzar garanteixen l'existència de la distribució estacionària $\mathbf{X}$. Per simplificar la seva resolució, veiem que el problema de minimització anterior es pot reescriure com un problema resoluble mitjançant programació lineal.

Es considera

$$\min_{\lambda} w$$

amb les condicions que

$$\begin{bmatrix} w \\ w \\ \vdots \\ w \end{bmatrix} \geq \hat{\mathbf{X}} - \begin{bmatrix} \hat{Q}_1\hat{\mathbf{X}} \cdot \lambda \\ \hat{Q}_2\hat{\mathbf{X}} \cdot \lambda \\ \vdots \\ \hat{Q}_n\hat{\mathbf{X}} \cdot \lambda \end{bmatrix}$$

$$\begin{bmatrix} w \\ w \\ \vdots \\ w \end{bmatrix} \geq -\hat{\mathbf{X}} + \begin{bmatrix} \hat{Q}_1\hat{\mathbf{X}} \cdot \lambda \\ \hat{Q}_2\hat{\mathbf{X}} \cdot \lambda \\ \vdots \\ \hat{Q}_n\hat{\mathbf{X}} \cdot \lambda \end{bmatrix}$$

$$w \geq 0 \ , \ \sum_{i=1}^{k} \lambda_i = 1 \ , \ \text{ i } \lambda_i \geq 0, \forall i$$

Notem que $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)^T$.

Utilitzant aquest mètode, els valors dels paràmetres λ_i es calculen d'una manera molt eficient.

Per la resolució d'aquest problema d'optimització hem utilitzat la norma $||\cdot||_\infty$. En canvi, si es fes servir la norma $||\cdot||_1$ el problema d'optimització es veuria modificat al següent:

$$\min_{\lambda} \left\{ \sum_{l=1}^{m} \left| \left[\sum_{i=1}^{k} \lambda_i \hat{Q}_i \hat{\mathbf{X}} - \hat{\mathbf{X}} \right]_l \right| \right\}$$

amb les condicions que

$$\sum_{i=1}^{k} \lambda_i = 1 \ \text{ i } \ \lambda_i \geq 0, \ \forall i$$

Per aquest cas, el problema de programació lineal corresponent serà

$$\min_{\lambda} \sum_{l=1}^{m} w_l$$

amb les condicions que

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \geq \hat{\mathbf{X}} - \begin{bmatrix} \hat{Q}_1\hat{\mathbf{X}} \cdot \lambda \\ \hat{Q}_2\hat{\mathbf{X}} \cdot \lambda \\ \vdots \\ \hat{Q}_n\hat{\mathbf{X}} \cdot \lambda \end{bmatrix}$$

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} \geq -\hat{\mathbf{X}} + \begin{bmatrix} \hat{Q}_1\hat{\mathbf{X}} \cdot \lambda \\ \hat{Q}_2\hat{\mathbf{X}} \cdot \lambda \\ \vdots \\ \hat{Q}_n\hat{\mathbf{X}} \cdot \lambda \end{bmatrix}$$

$$w_i \geq 0 \ , \ \sum_{i=1}^{k} \lambda_i = 1 \ , \ \text{i } \lambda_i \geq 0 \ ,\hat{\text{E}}\forall i$$

Notem que de nou $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)^T$.

4.3 Exemple

A continuació treballarem amb una seqüència $\{X^{(n)}\}$ de tres estats, $m = 3$,

$$\{X^{(n)}\} = \{2, 2, 3, 1, 2, 1, 1, 3, 1, 3, 2, 3, 2, 1, 3, 1, 1, 2, 1, 3, 3, 1, 2, 1, 2\}$$

Aquesta seqüència es pot escriure també de forma vectorial,

$$X^{(1)} = (0, 1, 0)^T, X^{(2)} = (0, 1, 0)^T, X^{(3)} = (0, 0, 1)^T, \ldots, X^{(24)} = (1, 0, 0)^T, X^{(25)} = (0, 1, 0)^T$$

Suposarem que es tracta d'una cadena de Markov d'ordre 2, aleshores les matrius de freqüència de transició seran

$$\Phi^{(1)} = \begin{bmatrix} 2 & 4 & 4 \\ 4 & 1 & 2 \\ 4 & 2 & 1 \end{bmatrix} \ \text{ i } \ \Phi^{(2)} = \begin{bmatrix} 5 & 2 & 3 \\ 2 & 2 & 2 \\ 2 & 3 & 2 \end{bmatrix}$$

Així doncs, les matrius de probabilitat de transició del pas i-èssim seran

$$\hat{Q}_1 = \begin{bmatrix} \frac{1}{5} & \frac{4}{7} & \frac{4}{7} \\ \frac{2}{5} & \frac{1}{7} & \frac{2}{7} \\ \frac{2}{5} & \frac{2}{7} & \frac{1}{7} \end{bmatrix} \ \text{ i } \ \hat{Q}_2 = \begin{bmatrix} \frac{5}{9} & \frac{2}{7} & \frac{3}{7} \\ \frac{2}{9} & \frac{2}{7} & \frac{2}{7} \\ \frac{2}{9} & \frac{3}{7} & \frac{2}{7} \end{bmatrix}$$

Efectivament, es constata que la suma per columnes de les matrius de probabilitat de transició del pas i-èssim és 1.

Per crear el vector $\hat{\mathbf{X}}$, en la seva component n-èssima escriurem la probabilitat d'aparició de l'estat n dintre la seqüència. En aquest cas, tindrem,

$$\hat{\mathbf{X}} = \left(\frac{2}{5}, \frac{8}{25}, \frac{7}{25}\right)^T$$

Per tant, es complirà que

$$\hat{Q}_1\hat{\mathbf{X}} = \left(\frac{74}{175}, \frac{2}{7}, \frac{51}{175}\right)^T$$

i

$$\hat{Q}_2\hat{\mathbf{X}} = \left(\frac{683}{1575}, \frac{82}{315}, \frac{482}{1575}\right)^T$$

Per tal d'estimar λ_i considerarem la norma $||\cdot||_\infty$ i tindrem el problema d'optimització següent

$$\min_{\lambda_1,\lambda_2} w$$

amb les condicions que

$$\begin{cases} w \geq \frac{2}{5} - \frac{74}{175}\lambda_1 - \frac{683}{1575}\lambda_2 \\ w \geq -\frac{2}{5} + \frac{74}{175}\lambda_1 + \frac{683}{1575}\lambda_2 \\ w \geq \frac{8}{25} - \frac{2}{7}\lambda_1 - \frac{82}{315}\lambda_2 \\ w \geq -\frac{8}{25} + \frac{2}{7}\lambda_1 + \frac{82}{315}\lambda_2 \\ w \geq \frac{7}{25} - \frac{51}{175}\lambda_1 - \frac{482}{1575}\lambda_2 \\ w \geq -\frac{7}{25} + \frac{51}{175}\lambda_1 + \frac{482}{1575}\lambda_2 \\ w \geq 0 \\ \lambda_1 + \lambda_2 = 1 \\ \lambda_1, \lambda_2 \geq 0 \end{cases}$$

Programant aquest problema lineal en MATLAB © s'arriba a la solució òptima següent:

$$(w^*, \lambda_1^*, \lambda_2^*) = (0.0343, 1, 0)$$

Per tant, el model queda de la forma següent:

$$\mathbf{X}^{(n+1)} = \hat{Q}_1\mathbf{X}^{(n)}$$

Cal notar la importància d'imposar que tant λ_1 com λ_2 siguin positius. En cas que no hi hagués aquesta condició, la solució òptima del problema seria:

$$(w^{**}, \lambda_1^{**}, \lambda_2^{**}) = (0.0083, 2.3500, -1.3500)$$

i el model quedaria com:

$$\mathbf{X}^{(n+1)} = 2.35\hat{Q}_1\mathbf{X}^{(n)} - 1.35\hat{Q}_2\mathbf{X}^{(n-1)}$$

Tot i que $w^{**} < w^*$, aquest darrer model no és correcte. Això es pot comprovar prenent, per exemple, $\mathbf{X}^{(n)} = \mathbf{X}^{(n-1)} = (1, 0, 0)^T$ i aleshores:

$$\mathbf{X}^{(n+1)} = (-0.28, 0.64, 0.64)^T$$

Com $\mathbf{X}^{(n+1)}$ és una distribució de probabilitat, el fet que hi hagi un valor negatiu clarament mostra que aquest darrer model és erroni.

D'una manera alternativa, si treballem amb la norma $||\cdot||_1$, tindrem el problema de minimització següent

$$\min_{\lambda_1, \lambda_2}(w_1 + w_2 + w_3)$$

amb les condicions que

$$\begin{cases} w_1 \geq \frac{2}{5} - \frac{74}{175}\lambda_1 - \frac{683}{1575}\lambda_2 \\ w_1 \geq -\frac{2}{5} + \frac{74}{175}\lambda_1 + \frac{683}{1575}\lambda_2 \\ w_2 \geq \frac{8}{25} - \frac{2}{7}\lambda_1 - \frac{82}{315}\lambda_2 \\ w_2 \geq -\frac{8}{25} + \frac{2}{7}\lambda_1 + \frac{82}{315}\lambda_2 \\ w_3 \geq \frac{7}{25} - \frac{51}{175}\lambda_1 - \frac{482}{1575}\lambda_2 \\ w_3 \geq -\frac{7}{25} + \frac{51}{175}\lambda_1 + \frac{482}{1575}\lambda_2 \\ w_1, w_2, w_3 \geq 0 \\ \lambda_1 + \lambda_2 = 1 \\ \lambda_1, \lambda_2 \geq 0 \end{cases}$$

i de nou programant en MATLAB, s'arriba al mateix resultat que abans,

$$(w_1^*, w_2^*, w_3^*, \lambda_1^*, \lambda_2^*) = (0.0229, 0.0343, 0.0114, 1, 0)$$

I el model tornarà a ser de la forma

$$\mathbf{X}^{(n+1)} = \hat{Q}_1\mathbf{X}^{(n)}$$

5 Aplicacions en la Neurofísica

5.1 Entropia

Definició 5.1.1. A la termodinàmica, l'**entropia** és una magnitud física que per un sistema termodinàmic en equilibri mesura el número de microestats compatibles amb el macroestat d'equilibri. També es pot dir que mesura el grau de desordre del sistema.

Sigui S l'entropia, k la constant de Boltzmann i Ω el número de microestats possibles pel sistema. Aleshores en la mecànica estadística es defineix,

$$S = k \ln \Omega$$

D'aquesta fórmula s'extreu que si només es pot arribar al macroestat d'una sola manera determinada, és a dir que ve descrit per un sol microestat, l'entropia serà nul·la. En canvi, si un mateix macroestat es pot definir mitjançant moltes combinacions possibles, és a dir per molts microestats, el valor de Ω serà alt i en conseqüència també el de l'entropia. És per això que es diu que l'entropia mesura el nivell de desordre d'un sistema.

Definició 5.1.2. Dintre de la teoria de la informació, l'**entropia**, també anomenada **entropia de Shannon**, mesura la incertesa d'una font d'informació. També es pot entendre com la quantitat d'informació promig que conté un símbol en un senyal enviat.

La idea és que aquesta entropia satisfaci que:

1. la mesura d'informació ha de ser proporcional, és a dir, que un canvi petit en una de les probabilitats d'aparició en uns dels elements del senyal ha de canviar poc l'entropia.
2. si tots els elements del senyal tenen la mateixa probabilitat d'aparèixer, aleshores l'entropia serà màxima.

Aquest darrer punt es deu a que en aquesta situació el nivell de desordre possible serà màxim, és a dir, màxima entropia, perquè no es podrà predir de cap manera el següent element del senyal ja que tots són equiprobables.

Sigui x_i un estat del procés X_i amb distribució de probabilitat p. Sigui A el conjunt dels estats possibles del procés X_i i sigui $\log \frac{1}{p(x_i)}$ la incertesa d'ocupar l'estat x_i. Aleshores, l'**entropia de Shannon** es defineix com,

$$H(X_i) = \sum_{x_i \in A} p(x_i) \log \frac{1}{p(x_i)}$$

Observació 5.1.1. Notem que com treballem amb probabilitats se satisfà que $0 \leq p(x_i) \leq 1$ i en conseqüència es trindrà que $H(X_i) \geq 0$.

Estudiem el cas en què tenim una assumpció a priori de la distribució de probabilitat del procés X_i que denotem per $q(x_i)$. Es quantificarà l'error que es fa quan s'utilitza $q(x_i)$ en comptes de la distribució de probabilitat real $p(x_i)$ mitjançant la diferència entre les incerteses, $\log \frac{1}{q(x_i)} - \log \frac{1}{p(x_i)}$. D'aquesta situació s'arriba a l'**entropia de Kullback**, que es defineix com:

$$K_{p|q}(X_i) = \sum_{x_i \in A} p(x_i) \log \frac{p(x_i)}{q(x_i)}$$

Observació 5.1.2. Veiem que efectivament $K_{p|q}(X_i) > 0$. Això es pot comprovar mitjançant la *desigualtat de Jensen*, que diu que per números arbitraris no negatius $a_1, \ldots, a_n$ i $b_1, \ldots, b_n$ es compleix

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

Per tant, es veu que

$$K_{p|q}(X_i) \geq \left(\sum p(x_i) \right) \log \frac{\sum p(x_i)}{\sum q(x_i)} = 0$$

A contuniació analitzem el cas particular d'entropia de Kullback més conegut i amb més repercussió. Suposem que treballem amb dos processos X_i i Y_j. Si fossin independents, la seva distribució de probabilitat conjunta factoritzaria. Així doncs, prenem l'assumpció a priori de l'entropia de Kullback que $q(x_i, y_j) = p(x_i)p(y_j)$, que dóna lloc a la **informació mútua** que es defineix com:

$$M(X_i, Y_j) = \sum_{x_i \in A} \sum_{y_j \in A} \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

Observació 5.1.3. Notem que en general, se satisfaria que $x_i \in A$ i que $y_j \in B$. En aquest cas més particular però, al tenir una aplicació en la física i al fer-se les mesures sobre el mateix sistema es dóna que $A = B$.

Cal notar que la informació mútua és una mesura de gran importància en moltes àrees de la física. Al tractar-se d'un cas particular d'entropia de Kullback, el que retorna és l'error comès en suposar que l'assumpció a priori és correcta. Així doncs, mitjançant la informació mútua es pot conéixer el grau de dependència que hi ha entre dos processos. Clarament, si $M(X_i, Y_j) = 0$, aquests dos processos seran independents.

Cal fer un èmfasi especial en que la informació mútua és un operador simètric.

A continuació tractarem les entropies dinàmiques. La principal diferència, és que l'estructura dinàmica d'un procés aleatori es reflexa per les probabilitats de transició.

Definició 5.1.3. El vector $\mathbf{X}_i^{(k)}$ té per components la composició dels processos dels k instants anteriors a i. $\mathbf{x}_i^{(k)}$ serà l'estat k-dimensional del procés $\mathbf{X}_i^{(k)}$.

En aquest cas de les entropies dinàmiques, la incertesa de la transició a un nou estat, donat els estats passats k-èssims, es defineix per $\log \frac{1}{p(x_{i+1}|\mathbf{x}_i^{(k)})}$. Mitjançant aquest valor d'incertesa, les entropies de la secció anterior es podem reescriure.

Prenent l'expressió de l'entropia de Shannon i fent la mitjana sobre tot els valors dels estats inicials $x_i^{(k)}$, es té que l'**entropia condicional de Shannon** és de la forma:

$$H(X_{i+1}|\mathbf{X}_i^{(k)}) = \sum_{\mathbf{x}_i^{(k)} \in A^k} p(\mathbf{x}_i^{(k)}) \sum_{x_{i+1} \in A} p(x_{i+1}|\mathbf{x}_i^{(k)}) \log \frac{1}{p(x_{i+1}|\mathbf{x}_i^{(k)})} =$$

$$= \sum_{x_{i+1}, \mathbf{x}_i^{(k)}} p(x_{i+1}, \mathbf{x}_i^{(k)}) \log \frac{1}{p(x_{i+1}|\mathbf{x}_i^{(k)})}$$

Normalment, quan s'estudia un procés amb una dinàmica intrínseca desconeguda, s'ha d'assumir una probabilitat de transició a priori en comptes de la probabilitat de transició autèntica. Així doncs, de manera equivalent al cas estàtic, es pot construir l'**entropia de Kullback condicionada** que és de la forma,

$$K_{p|q}(X_{i+1}|\mathbf{X}_i^{(k)}) = \sum_{x_{i+1}, \mathbf{x}_i^{(k)}} p(x_{i+1}, \mathbf{x}_i^{(k)}) \log \frac{p(x_{i+1}|\mathbf{x}_i^{(k)})}{q(x_{i+1}|\mathbf{x}_i^{(k)})}$$

De manera equivalent al que s'ha vist abans, agafant l'assumpció a prori de l'entropia de Kullback condicionada que els processos dinàmics $\mathbf{X}_i^{(k)}$ i $\mathbf{Y}_j^{(l)}$ són independents, s'arriba a la **informació mútua condicionada** que té la forma:

$$M(X_{i+1}, Y_{j+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) = \sum_{x_{i+1}, \mathbf{x}_i^{(k)}} \sum_{y_{j+1}, \mathbf{y}_j^{(l)}} p(x_{i+1}, \mathbf{x}_i^{(k)}, y_{j+1}, \mathbf{y}_j^{(l)}) \log \frac{p(x_{i+1}, y_{j+1}, \mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)})}{p(x_{i+1}, \mathbf{x}_i^{(k)}) p(y_{j+1}, \mathbf{y}_j^{(l)})}$$

De nou, s'aprecia que aquesta mesura és simètrica. Així doncs, sí que quantificarà la dependència dinàmica entre ambdós processos però això no és suficient. Per l'aplicació a la neurofísica, on s'estudien les connexions entre neurones, no n'hi ha prou en saber quant dependents són els senyals elèctrics de dues neurones, sinó que també es necessita conèixer la direcció en la qual es produeix l'intercanvi d'informació. Per entendre el comportament d'un cultiu neuronal no n'hi ha prou en saber com de connectades estan les neurones A i B, sinó que cal saber si la neurona A excita a la B o viceversa. És per això que es va desenvolupar l'eina matemàtica següent.

Definició 5.1.4. La **transferència d'entropia** T es defineix com:

$$T(X_{i+1}|\mathbf{X}_i^{(k)}, \mathbf{Y}_j^{(l)}) = \sum_{x_{i+1}, \mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)}} p(x_{i+1}, \mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)}) \log \frac{p(x_{i+1}|\mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)})}{p(x_{i+1}|\mathbf{x}_i^{(k)})}$$

Cal notar diferents propietats d'aquesta mesura. Per començar, cal recalcar que és un cas particular de l'entropia de Kullback condicionada i per tant T és no negativa. Per altra banda, la transferència d'entropia no és simètrica sota intercanvi de X i Y. Això permetrà que tingui en compte la direccionalitat de la transferència d'informació. A més a més, en la transferència d'entropia es pren $p(x_{i+1}|\mathbf{x}_i^{(k)}, \mathbf{y}_j^{(l)})$ com la probabilitat de transició intrínseca del procés mentre que a prori s'assumeix que la probabilitat de transició és $p(x_{i+1}|\mathbf{x}_i^{(k)})$. Així doncs, la transferència d'entropia quantifica la dependència que el futur estat x_{i+1} del procés X_i depèn dels k estats anteriors de $\mathbf{X}_i^{(k)}$ però no dels l estats de $\mathbf{Y}_j^{(l)}$. D'una manera alternativa, es pot entendre com que, al ser un cas particular de l'entropia de Kullback condicionada, la transferència d'entropia mesura l'error en suposar que l'evolució del procés X només vindrà condicionada per ell mateix i no per l'altre procés Y. A part de la direccionalitat aporta informació de quant depenents són ambdós processos entre si.

L'exemple següent permet apreciar les diferències entre la informació mútua i la transferència d'entropia i el fet que aquesta darrera mesura és asimètrica.

5.2 Exemples

5.2.1 Exemple genèric

Es consideren dos processos de Markov X i Y d'estats binaris estacionaris. Es compleix que X és independent però Y depèn de l'estat de X. Per l'estat X es té que passa de l'estat 0 al 1 i de l'1 al 0 amb probabilitat 1 a cada pas. El procés depèn només de la condició inicial. Pel que fa a l'estat Y, depèn del X de la següent manera: la probabilitat que $y_{i+1} = x_i$ és $\frac{1+c}{2}$ i que $y_{i+1} = 1 - x_i$ és $\frac{1-c}{2}$, on c és una constant. No hi ha cap tipus de dependència entre y_{i+1} i y_i.

Per tant,

$$\begin{cases} p(y_{i+1} = 1|x_i = 1) = p(y_{i+1} = 0|x_i = 0) = \frac{1+c}{2} \\ p(y_{i+1} = 1|x_i = 0) = p(y_{i+1} = 0|x_i = 1) = \frac{1-c}{2} \end{cases}$$

Mitjançant aquestes regles es poden calcular les probabilitats condicionades $p(x_{i+1}, y_{i+1}|x_i, y_i)$. Com X és independent i no hi ha dependència entre y_{i+1} i y_i es té que:

$$\begin{cases} p(x_{i+1}=1|x_i=0)=p(x_{i+1}=0|x_i=1)=1 \\ p(x_{i+1}=1|x_i=1)=p(x_{i+1}=0|x_i=0)=0 \\ p(y_{i+1}=a|y_i=b)=p(y_{i+1}=c)=p(y_i=d)=\frac{1}{2} \quad \text{on } a,b,c,d\in\{0,1\} \end{cases}$$

I com a conseqüència també se satisfà que $p(x_{i+1},y_{i+1}|x_i,y_i)=p(x_{i+1},y_{i+1}|x_i)$, el que simplifica els càlculs. Al final es tindrà que:

$$p(x_{i+1},y_{i+1}|x_i,y_i)=p(x_{i+1},y_{i+1}|x_i)=\begin{cases} 0 & \text{si } x_{i+1}=x_i \\ \frac{1+c}{2} & \text{si } x_{i+1}\neq x_i \text{ i } y_{i+1}=x_i \\ \frac{1-c}{2} & \text{si } x_{i+1}\neq x_i \text{ i } y_{i+1}=1-x_i \end{cases}$$

Per altra banda, es compleix que:

$$\begin{cases} p(x_i=1,y_i=1)=p(x_i=0,y_i=0)=\frac{1+c}{4} \\ p(x_i=1,y_i=0)=p(x_i=0,y_i=1)=\frac{1-c}{4} \end{cases}$$

Així doncs, mitjançant la combinació d'aquestes dues es calcula $p(x_{i+1},y_{i+1},x_i,y_i)$. Prenent que $k=l=1$ procedim primer a calcular la informació mútua:

$$M(X_{i+1},Y_{i+1}|X_i,Y_i)=\sum_{x_{i+1},x_i}\sum_{y_{i+1},y_i}p(x_{i+1},x_i,y_{i+1},y_i)\log\frac{p(x_{i+1},y_{i+1}|x_i,y_i)}{p(x_{i+1}|x_i)p(y_{i+1}|y_i)}$$

Substituint tots els valors anteriors s'arriba a:

$$M(X_{i+1},Y_{i+1}|X_i,Y_i)=\log(1+c)\left[\frac{1+c}{2}\cdot\frac{1+c}{4}+\frac{1+c}{2}\cdot\frac{1-c}{4}+\frac{1+c}{2}\cdot\frac{1-c}{4}+\frac{1+c}{2}\cdot\frac{1+c}{4}\right]+$$

$$+\log(1-c)\left[\frac{1-c}{2}\cdot\frac{1-c}{4}+\frac{1-c}{2}\cdot\frac{1+c}{4}+\frac{1-c}{2}\cdot\frac{1+c}{4}+\frac{1-c}{2}\cdot\frac{1-c}{4}\right]\Rightarrow$$

$$M(X_{i+1},Y_{i+1}|X_i,Y_i)=\frac{1+c}{2}\log(1+c)+\frac{1-c}{2}\log(1-c)$$

A continuació procedim a calcular la transferència d'entropia per ambdós casos:

1. $T(X_{i+1}|X_i,Y_i)=\sum_{x_{i+1},x_i,y_i}p(x_{i+1},x_i,y_i)\log\frac{p(x_{i+1}|x_i,y_i)}{p(x_{i+1}|x_i)}$

   Notem que pel mateix argument anterior, que X és un procés independent, es complirà que:

$$p(x_{i+1}|x_i, y_i) = p(x_{i+1}|x_i)$$

Treballant només amb els casos possibles, és a dir, que el procés X passi de 1 a 0 o de 0 a 1, s'arribarà al valor log(1), el que provocarà que tot s'anul·li, i com a conseqüència:

$$T(X_{i+1}|X_i, Y_i) = 0$$

Analitzant aquest resultat veiem que conceptualment és correcte. El que ens està retornant el valor de la transferència d'entropia és l'error que es comet a l'assumir que l'estat futur x_{i+1} de X depèn de l'instant anterior del propi procés X i no de l'Y. Efectivament no es comet cap error al suposar això ja que el procés X és totalment independent de l'Y.

2. $T(Y_{i+1}|Y_i, X_i) = \sum_{y_{i+1},y_i,x_i} p(y_{i+1}, y_i, x_i) \log \frac{p(y_{i+1}|y_i, x_i)}{p(y_{i+1}|y_i)}$

   Com no hi ha cap tipus de dependència entre y_{i+1} i y_i aleshores es compleix que:

$$p(y_{i+1}|y_i, x_i) = p(y_{i+1}|x_i)$$

   que són probabiltats per les quals ja coneixem el valor, igual que per $p(y_{i+1}|y_i)$. Falta calcular el valor de $p(y_{i+1}, y_i, x_i)$, que es farà utilitzant:

$$p(y_{i+1}, y_i, x_i) = p(y_{i+1}|y_i, x_i) \cdot p(y_i, x_i) = p(y_{i+1}|x_i) \cdot p(x_i, y_i)$$

   Substituint els valors corresponents s'arriba a que:

$$\begin{array}{rr}
p(y_{i+1} = 0, y_i = 0, x_i = 0) = \frac{(1+c)^2}{8}, & p(y_{i+1} = 0, y_i = 0, x_i = 1) = \frac{(1-c)^2}{8} \\
p(y_{i+1} = 0, y_i = 1, x_i = 0) = \frac{1-c^2}{8}, & p(y_{i+1} = 1, y_i = 0, x_i = 0) = \frac{1-c^2}{8} \\
p(y_{i+1} = 0, y_i = 1, x_i = 1) = \frac{1-c^2}{8}, & p(y_{i+1} = 1, y_i = 0, x_i = 1) = \frac{1-c^2}{8} \\
p(y_{i+1} = 1, y_i = 1, x_i = 0) = \frac{(1-c)^2}{8}, & p(y_{i+1} = 1, y_i = 1, x_i = 1) = \frac{(1+c)^2}{8}
\end{array}$$

   Fent servir ara que $p(y_{i+1}, y_i) = p(y_{i+1}, y_i, x_i = 0) + p(y_{i+1}, y_i, x_i = 1)$, s'arriba a que:

$$\begin{array}{rr}
p(y_{i+1} = 0, y_i = 0) = \frac{1+c^2}{4}, & p(y_{i+1} = 0, y_i = 1) = \frac{1-c^2}{4} \\
p(y_{i+1} = 1, y_i = 0) = \frac{1-c^2}{4}, & p(y_{i+1} = 1, y_i = 1) = \frac{1+c^2}{4}
\end{array}$$

   Com $p(y_{i+1}|y_i) = \frac{p(y_{i+1},y_i)}{p(y_i)}$, es té que:

$$
\begin{array}{ll}
p(y_{i+1}=0|y_i=0)=\frac{1+c^2}{2}, & p(y_{i+1}=0|y_i=1)=\frac{1-c^2}{2} \\
p(y_{i+1}=1|y_i=0)=\frac{1-c^2}{2}, & p(y_{i+1}=1|y_i=1)=\frac{1+c^2}{2}
\end{array}
$$

$$
T(Y_{i+1}|Y_i,X_i)=\ldots=2\cdot\frac{1-c^2}{8}\log\frac{1+c}{(1+c)(1-c)}+2\cdot\frac{1-c^2}{8}\log\frac{1-c}{(1+c)(1-c)}+
$$

$$
+2\cdot\frac{(1+c)^2}{8}\log\frac{1+c}{1+c^2}+2\cdot\frac{(1-c)^2}{8}\log\frac{1-c}{1+c^2}
$$

$$
T(Y_{i+1}|Y_i,X_i)=\frac{1}{2}c[(1+c)\log(1+c)-(1-c)\log(1-c)]-\frac{1}{2}(1+c^2)\log(1+c^2)
$$

Observació 5.2.1. Amb aquest exemple queda patent que $T(X_{i+1}|X_i,Y_i)\neq T(Y_{i+1}|Y_i,X_i)$.

5.2.2 Exemple en la neurofísica

La neurofísica, entre d'altres coses, pretèn estudiar les connexions entre neurones. No és una tasca gens fàcil degut al número elevadíssim de connexions que es troben en el cervell. Per tal de poder fer aquest estudi més assequible, s'estudien cultius neuronals d'unes 100 neurones aproximadament. Un procediment per tal de poder extreure informació d'aquests cultius és injectar un pigment a les neurones que és reactiu amb els ions de calci. S'escull aquest pigment perquè hi ha un trasbàs d'ions de calci quan passa un senyal elèctric per una neurona. Així doncs, quan passi un senyal elèctric, les neurones esdevindran fluorescents degut al moviment dels ions de calci. Aquests senyals de fluorescència s'extreuen mitjançant la *calcuim imaging technique*. Amb aquestes dades s'inferirà el perfil del senyal elèctric de la neurona mitjançant el *peeling algorithm*. El resultat d'aquest algorisme es pot apreciar a la figura següent.

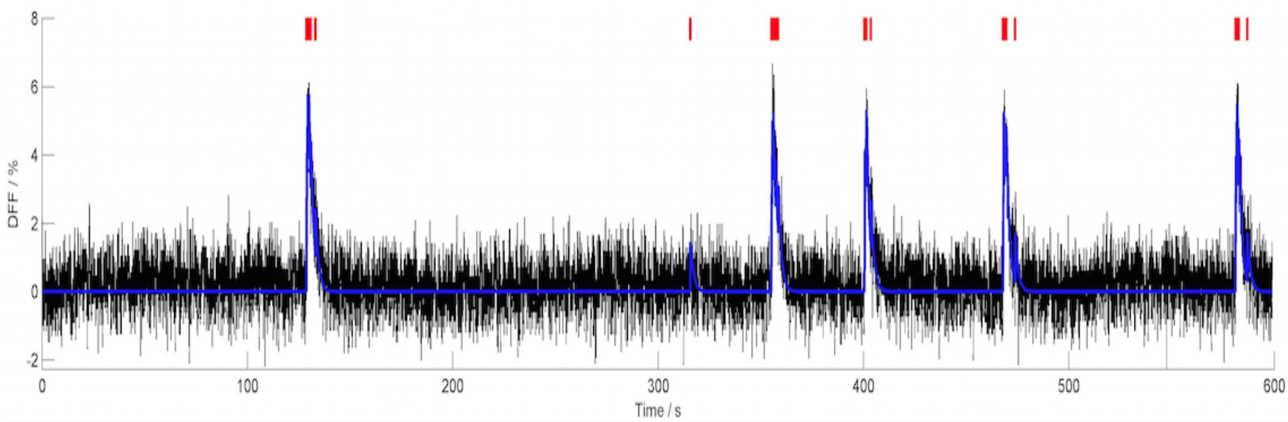

Figura 5: *Representació del senyal elèctric d'una neurona després d'aplicar el peeling algoritm*

En aquestes mesures hi ha molt de soroll, tal com es pot veure a la figura en les contínues fluctuacions de la traça negra. És per això que al *peeling algorith* se li ha de donar un valor llindar per tal de decidir si hi ha o no un pic en el senyal. És a dir, si l'amplitud del senyal supera el valor llindar es dirà que hi ha un pic i en cas contrari que no n'hi ha. Des d'un punt de vist biològic, aquest pic representa que en aquell moment la neurona ha enviat un senyal. Aquest senyal elèctric es tradueix en un senyal binari, on l'1 representa la presència d'un pic i el 0 l'absència d'un en el senyal. I és a partir d'aquest senyal binari que es calcularan els valors de la transferència d'entropia. Veiem a continuació un exemple del càlcul de la transferència d'entropia amb els senyals binaris de dues neurones, prenent $k = l = 1$.

Suposem que tenim el següent senyal binari de dues neurones X i Y.

$$X : 1110110100110011101101$$
$$Y : 1010101011011011011001$$

Procedim a calcular els valors de la transferència d'entropia, però abans cal comentar que tot i que els senyals tinguin una llargada de 22, només treballarem amb els 21 primers elements perquè com ens interessa calcular valors per x_{i+1}, l'últim element no ens serveix perquè no tenim informació del seu element posterior.

1. $T(X_{i+1}|X_i, Y_i) = \sum_{x_{i+1}, x_i, y_i} p(x_{i+1}, x_i, y_i) \log \frac{p(x_{i+1}|x_i, y_i)}{p(x_{i+1}|x_i)}$

   on es farà servir que:

$$p(x_{i+1}|x_i, y_i) = \frac{p(x_{i+1}, x_i, y_i)}{p(x_i, y_i)} \quad \text{i} \quad p(x_{i+1}|x_i) = \frac{p(x_{i+1}, x_i)}{p(x_i)}$$

   Les probabilitats prendran els següents valors:

$$\begin{aligned} p(x_{i+1} = 0, x_i = 0, y_i = 0) &= 0, & p(x_{i+1} = 0, x_i = 0, y_i = 1) &= \tfrac{2}{21} \\ p(x_{i+1} = 0, x_i = 1, y_i = 0) &= \tfrac{4}{21}, & p(x_{i+1} = 1, x_i = 0, y_i = 0) &= \tfrac{3}{21} \\ p(x_{i+1} = 0, x_i = 1, y_i = 1) &= \tfrac{2}{21}, & p(x_{i+1} = 1, x_i = 0, y_i = 1) &= \tfrac{3}{21} \\ p(x_{i+1} = 1, x_i = 1, y_i = 0) &= \tfrac{2}{21}, & p(x_{i+1} = 1, x_i = 1, y_i = 1) &= \tfrac{5}{21} \end{aligned}$$

$$\begin{aligned} p(x_i = 0, y_i = 0) &= \tfrac{3}{21}, & p(x_i = 0, y_i = 1) &= \tfrac{5}{21} \\ p(x_i = 1, y_i = 0) &= \tfrac{6}{21}, & p(x_i = 1, y_i = 1) &= \tfrac{7}{21} \end{aligned}$$

$$\begin{aligned} p(x_{i+1} = 0, x_i = 0) &= \tfrac{2}{21}, & p(x_{i+1} = 0, x_i = 1) &= \tfrac{6}{21} \\ p(x_{i+1} = 1, x_i = 0) &= \tfrac{6}{21}, & p(x_{i+1} = 1, x_i = 1) &= \tfrac{7}{21} \end{aligned}$$

$$p(x_i = 0) = \tfrac{8}{21}, \qquad p(x_i = 1) = \tfrac{13}{21}$$

Amb aquests valors ja es poden calcular les quantitats de $p(x_{i+1}|x_i, y_i)$ i de $p(x_{i+1}|x_i)$. Substituint-ho tot a la fórmula de transferència d'entropia s'arriba a que:

$$T(X_{i+1}|X_i, Y_i) = \ldots = 0.0434$$

2. $T(Y_{i+1}|Y_i, X_i) = \sum_{y_{i+1}, y_i, x_i} p(y_{i+1}, y_i, x_i) \log \frac{p(y_{i+1}|y_i, x_i)}{p(y_{i+1}|y_i)}$

   Repetint un procediment totalment equivalent al del cas anterior, calculant totes les probabilitats necessàries i substituint el resultat a la fórmula s'obté que en aquest cas:
$$T(Y_{i+1}|Y_i, X_i) = \ldots = 0.0108$$

Un cop s'han calculat els valors de transferència d'entropia, com es poden extreure conclusions de les connexions entre les neurones? El que es fa és crear un graf dirigit on els nodes representen les neurones del cultiu i les arestes les connexions. Recordant la definició de transferència d'entropia, aquesta el que retorna és l'error en suposar que l'esdeveniment futur del procés X només dependrà d'ell mateix i no del procés Y també. Així doncs, si el valor de la transferència d'entropia és major, vol dir que l'error que s'ha comès al suposar això és més gran. Per tant, quant més gran sigui l'error, més connectats estaran els dos processos. En conseqüència, a les arestes del graf dirigit se'ls hi donarà com a pes el valor de la transferència d'entropia. D'aquesta manera, es pot estudiar la connectivitat del cultiu. Com la transferència d'entropia és asimètrica, el graf serà dirigit i les arestes prendran diferents pesos: no serà igual la que va del node A al B que la que va del B a l'A i així es pot saber si la connexió entre neurones és forta i també quina excita a quina.

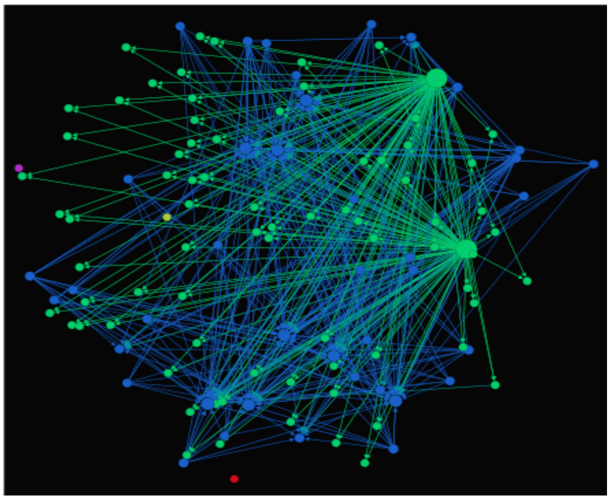

Figura 6: *Graf dirigit d'un cultiu neuronal per un valor llindar donat*

6 Conclusions

L'objectiu d'aquest treball era desenvolupar el temari necessari relacionat amb les cadenes de Markov i les cadenes de Markov d'ordre superior per tal de poder comprendre el funcionament de l'eina matemàtica anomenada *transferència d'entropia*. Mitjançant aquesta, es pot fer una anàlisi més detallada de cultius neuronals per entendre millor el comportament de les neurones i les relacions entre elles. La *transferència d'entropia* ha estat un avenç en el camp de la neurofísica perquè quantifica la intensitat de la relació entre dues neurones i també la direccionalitat en la qual emeten els senyals elèctrics.

No obstant, encara s'ha de refinar la tècnica per l'estudi dels cultius neuronals, dels que després s'extreu coneixement per comprendre el cervell. Per una banda, s'ha d'escollir un valor llindar per a obtenir el senyal elèctric que després es traduirà en un senyal binari amb el qual es calcularà la *transferència d'entropia*. Depenent del valor llindar escollit s'obtindran valors diferents sobre el comportament del cultiu. Així doncs, seria òptim desenvolupar un sistema que no fos tan sensible al valor d'un paràmetre que s'ha d'escollir a priori. Per altra banda, a l'hora de calcular la *transferència d'entropia* per cadenes de Markov d'ordre superior la complexitat computacional es dispara. És per això que tot i que els resultats que dóna aquesta eina són molt bons, un model on la complexitat computacional per cadenes de Markov fos encara menor agilitzaria el procés millorant l'estudi dels cultius.

Referències

[1] Karlin, S.; Taylor, H. M.: *A first course in stochastic processes*, 2a edició, Academic Press, 1975.

[2] Karlin, S.; Taylor, H.M.: *A second course in stochastic processes*, 1a edició, Academic Press, 1981.

[3] Norris, J.R.: *Markov chains*, 1a edició, Cambridge University Press, 1997.

[4] Chung, K.L.: *Markov chains with ctationary transition probabilities*, 2a edició, Springer, 1967.

[5] çinlar, E.: *Introduction to stochastic processes*, 1a edició, Prentice-Hall, 1975.

[6] Ash, R.B.: Basic probability theory, 1a edició, Wiley, 1970.

[7] Ching, W.; Huang, X.; Ng, M.K.; Siu, T.: *Markov Chains: Models, Algorithms and Applications*, 2a edició, Springer, 2013.

[8] Schreiber, T.; Kaiser, A.: Information transfer in continuous processes, *Physica D*, 166:43-62, 2002.