Capitolo 13

CATENE DI MARKOV

Le catene di Markov sono sistemi dinamici a tempo discreto che presentano caratteristiche alquanto diverse rispetto a quelli considerati finora. Nell'accezione più semplice, si tratta di sistemi autonomi a stati finiti[1] nei quali - e ciò rappresenta una essenziale novità rispetto ai modelli analizzati finora - la transizione da uno stato all'altro avviene su base probabilistica, anziché deterministica. In altre parole, l'informazione disponibile circa una catena al generico istante t è fornita dalle probabilità che essa si trovi in uno qualsiasi degli stati, e l'evoluzione temporale della catena viene specificata precisando in qual modo tali probabilità si aggiornino passando dall'istante t all'istante $t+1$. La trattazione delle catene di Markov si riconduce a quella dei sistemi discreti positivi quando si interpreti il comportamento dinamico di una catena in termini di evoluzione temporale dei vettori che esprimono la distribuzione di probabilità sui suoi vari stati. Tale evoluzione si ottiene infatti trasformando i vettori di probabilità mediante l'azione di matrici positive di una classe particolare (matrici stocastiche)[2].

13.1 Definizioni e concetti introduttivi

Una catena di Markov (finita) $\mathcal{C}$ fornisce un modello teorico per descrivere il comportamento di un sistema a tempo discreto che può trovarsi ad ogni istante in qualche stato appartenente ad un dato insieme finito $\mathbf{S} = \{S_1, S_2, \dots, S_n\}$. Passando da un istante al successivo, il sistema "salta" da uno stato ad un altro e la legge di carattere probabilistico secondo cui avviene la transizione è formulabile in questi termini:

> "qualunque sia stata l'evoluzione di $\mathcal{C}$ prima dell'istante t, se $\mathcal{C}$ al tempo t si trova nello stato S_i, all'istante $t+1$ si troverà nello stato S_j con probabilità p_{ij}".

L'insieme delle probabilità di transizione $\{p_{ij},\ i,j = 1,2,\dots,n\}$ è assegnato e determina il comportamento (probabilistico) della catena, una volta che siano note le condizioni iniziali (i.e. qual è al tempo $t=0$ lo stato in cui la catena si trova o, più in generale, qual è al tempo $t=0$ la distribuzione di probabilità sui vari stati della catena).

[1] Non considereremo le catene di Markov con un'infinità numerabile di stati, per la cui trattazione sono necessari strumenti analitici più complessi.

[2] Nella letteratura sulle catene di Markov è consuetudine (alla quale ci atterremo) rappresentare i vettori di probabilità come vettori riga e conseguentemente ottenerne l'aggiornamento temporale applicando le matrici stocastiche alla loro destra.

Osservazione In termini più formali, una catena di Markov (finita) $\mathcal{C}$ consiste di

- un insieme $\mathbf{S} = \{S_1, S_2, \dots, S_n\}$, chiamato "spazio di stato";
- una sequenza di variabili casuali $X(0), X(1), X(2), \dots$ ciascuna delle quali prende valori in $\mathbf{S}$ (il valore di $X(t)$ è lo stato che la catena "visita" nell'istante t), e aventi la seguente "proprietà di markovianità":

  se[3] $\operatorname{pr}\left\{X(t) = S_{\nu_t}, X(t-1) = S_{\nu_{t-1}}, \dots, X(0) = S_{\nu_0}\right\} > 0$, allora

$$\begin{aligned} & \operatorname{pr}\left\{X(t+1) = S_{\nu_{t+1}} \mid X(t) = S_{\nu_t}, X(t-1) = S_{\nu_{t-1}}, \dots, X(0) = S_{\nu_0}\right\} \\ = \; & \operatorname{pr}\left\{X(t+1) = S_{\nu_{t+1}} \mid X(t) = S_{\nu_t}\right\} \end{aligned}$$

La markovianità equivale ad asserire che l'evoluzione delle probabilità future (i.e. la distribuzione di probabilità al tempo $t+1$ sui possibili valori che $X(t+1)$ può assumere) è determinata dalla conoscenza delle condizioni attuali della catena (i.e. dalla conoscenza di quale valore assume $X(t)$), indipendentemente dal comportamento del processo (i.e. da quali stati ha visitato) negli istanti precedenti t. In questi appunti assumeremo che le probabilità $\operatorname{pr}\{X(t+1) = S_j | X(t) = S_i\}$ siano indipendenti[4] da t: si tratta appunto delle probabilità di transizione costanti p_{ij} di cui si è detto prima.

Le probabilità di transizione vengono raggruppate nella *matrice di transizione* della catena

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ & \cdots & \cdots & \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix},$$

una matrice positiva nella quale la somma degli elementi di ciascuna riga è unitaria. Infatti, gli elementi della riga i-esima di P sono le probabilità che la catena, trovandosi nello stato S_i all'istante t, transiti in S_1 o in $S_2, \dots$ o in S_n all'istante successivo; poiché tali eventi di transizione risultano mutuamente esclusivi ed esaustivi di tutte le possibilità, la somma delle loro probabilità vale 1. Una matrice siffatta (positiva, quadrata, con righe a somma unitaria) si dirà "stocastica", e chiameremo "stocastico" ogni vettore riga positivo $\mathbf{x}^T = [\, x_1 \quad x_2 \quad \dots \quad x_n \,]$ avente somma degli elementi unitaria: $\sum_{i=1}^n x_i = 1$.

In alternativa alla matrice di transizione, la catena $\mathcal{C}$ può essere rappresentata mediante un *grafo orientato etichettato* nel quale

- i vertici corrispondono agli stati $S_1, S_2, \dots, S_n$ (o agli indici $1, 2, \dots, n$ degli stati);
- c'è un arco orientato che connette il vertice S_i al vertice S_j se e solo se è positiva la probabilità p_{ij} di transizione da S_i a S_j e tale probabilità viene utilizzata come "etichetta" dell'arco stesso.

Matrice di transizione e grafo forniscono la medesima informazione sulla catena $\mathcal{C}$.

Esempio 13.1.1 [Condizioni meteorologiche sulla vetta di Utopia] Nel paese di Utopia c'è una sola vetta degna di questo nome. Su essa non ci sono mai due giorni consecutivi di bel tempo, e quando non è bello o nevica o tempesta. Le pazienti osservazioni dei meteorologi hanno portato alle seguenti conclusioni:

- se un giorno c'è sole, il giorno dopo ci sono eguali probabilità che sulla vetta tempesti o nevichi;
- se tempesta o nevica, per il giorno successivo ci sono eguali probabilità che il tempo sulla vetta rimanga lo stesso, oppure che il tempo cambi;
- se tempesta o nevica e il tempo cambia, le probabilità che il giorno successivo la vetta sia in pieno sole o che la tempesta (la neve) del giorno prima si volga in neve (in tempesta) sono uguali.

[3] se $\operatorname{pr}\{B\} = 0$, $\operatorname{pr}\{A|B\}$ non è definita

[4] se le probabilità di transizione non dipendono da t, la catena si dice "omogenea"

Indicando nell'ordine con S, T, N (o con $1, 2, 3$) le tre possibili condizioni meteorologiche (sole, tempesta, neve), si ricavano la matrice di transizione

$$P = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

e il grafo orientato corrispondente di figura 13.1.1.

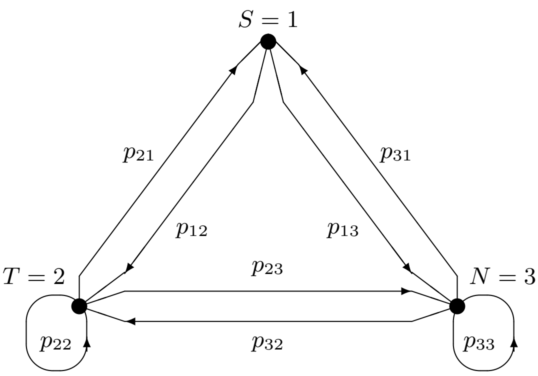


Figura 13.1.1

Esempio 13.1.2 [Una catena di Markov?] Per uno studente che frequenta un corso di laurea triennale, gli stati possibili della carriera sono:

- S_1 essere iscritto al primo anno di corso
- S_2 essere iscritto al secondo anno di corso
- S_3 essere iscritto al terzo anno di corso
- S_4 essere fuori corso
- S_5 conseguire la laurea
- S_6 ritirarsi dal corso di laurea

Trascorso un anno dall'iscrizione al primo anno di corso, egli ha probabilità $0,60$ di iscriversi al secondo, $0,20$ di ritirarsi, $0,20$ di ripetere l'iscrizione al primo anno;

trascorso un anno dall'iscrizione al secondo anno di corso, con probabilità $0,80$ si iscriverà al terzo anno, con probabilità $0,05$ si ritirerà, con probabilità $0,15$ ripeterà l'iscrizione al secondo anno;

trascorso un anno dall'iscrizione al terzo anno di corso, con probabilità $0,7$ si laureerà, con probabilità $0,02$ si ritirerà, con probabilità $0,28$ diventerà fuori corso;

alla conclusione di un anno di fuori corso, con probabilità $0,6$ conseguirà la laurea, con probabilità $0,38$ si reiscriverà come fuori corso, con probabilità $0,02$ si ritirerà dagli studi.

Con riferimento alla numerazione degli stati che abbiamo scelto, ricaviamo la matrice

$$\begin{array}{rc} & \begin{array}{cccccc} \underline{\text{a}}\ \ \text{anno I} & \text{II} & \text{III} & \text{FC} & \text{Lau} & \text{Rit} \end{array} \\ P \;=\; \begin{array}{r} \underline{\text{da}}\ \ \text{anno I} \\ \text{anno II} \\ \text{anno III} \\ \text{fuoricorso} \\ \text{laureato} \\ \text{ritirato} \end{array} & \begin{bmatrix} 0,20 & 0,60 & 0 & 0 & 0 & 0,20 \\ 0 & 0,15 & 0,80 & 0 & 0 & 0,05 \\ 0 & 0 & 0 & 0,28 & 0,70 & 0,02 \\ 0 & 0 & 0 & 0,38 & 0,60 & 0,02 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

Quali obiezioni possono essere mosse all'affermazione che l'evoluzione del vettore di probabilità abbia caratteristiche markoviane? Ad esempio, è corretto ritenere che la probabilità di reiscriversi al primo anno sia la medesima per uno studente che si è immatricolato da un anno e per quello che si è immatricolato da due?

Se la catena al tempo 0 si trova (deterministicamente) nello stato S_i, al tempo 1 potrà trovarsi in ciascuno degli stati $S_j,\ j = 1, 2, \ldots, n$ con probabilità p_{ij}. Il vettore stocastico

$\mathbf{x}^T(1)$, che fornisce la distribuzione di probabilità sugli stati di $\mathcal{C}$ all'istante $t = 1$,

$$\mathbf{x}^T(1) = [\, p_{i1} \quad p_{i2} \quad \dots \quad p_{in} \,]$$

si ottiene moltiplicando $\mathbf{e}_i^T$, che possiamo interpretare come distribuzione di probabilità sugli stati di $\mathcal{C}$ all'istante $t = 0$, per la matrice di transizione:

$$\mathbf{x}^T(1) = \mathbf{e}_i^T P$$

Se lo stato al tempo 0 è noto solo in termini probabilistici, ovvero se l'informazione disponibile è costituita dalle probabilità π_j che lo stato del sistema sia $S_j,\ j = 1, 2, \dots, n$, posto $\mathbf{x}^T(0) = [\, \pi_1 \quad \pi_2 \quad \dots \quad \pi_n \,]$, abbiamo[5]

$$\mathbf{x}^T(1) = [\, \pi_1 \quad \pi_2 \quad \dots \quad \pi_n \,]\, P. \tag{13.1}$$

Infatti, l'evento che la catena $\mathcal{C}$ al tempo $t = 1$ si trovi in S_j può verificarsi secondo le seguenti modalità, mutuamente esclusive ed esaustive di tutte le possibilità:

- per $t = 0$ lo stato è S_1 e si ha la transizione $S_1 \to S_j$,
- per $t = 0$ lo stato è S_2 e si ha la transizione $S_2 \to S_j$,

  ...
- per $t = 0$ lo stato è S_n e si ha la transizione $S_n \to S_j$.

La prima eventualità ha probabilità $\pi_1 p_{1j}$, la seconda ha probabilità $\pi_2 p_{2j}$, ..., l'ultima ha probabilità $\pi_n p_{nj}$, da cui la (13.1): il prodotto di $\mathbf{x}^T(0)$ per la j-esima colonna di P dà quindi la probabilità che lo stato della catena all'istante $t = 1$ sia S_j.
Un ragionamento analogo consente di determinare la distribuzione di probabilità al generico istante $t \geq 1$: tramite la (13.1), si esprime in funzione di $\mathbf{x}^T(t-1)$ il vettore stocastico $\mathbf{x}^T(t)$, le cui componenti sono le probabilità che la catena al tempo t si trovi in uno degli n stati possibili; a sua volta si esprime il vettore stocastico $\mathbf{x}^T(t-1)$ in funzione di $\mathbf{x}^T(t-2)$; etc. .... In tal modo si perviene alla relazione

$$\mathbf{x}^T(t) = \mathbf{x}^T(t-1)P = \left(\mathbf{x}^T(t-2)P\right) P = \dots = \mathbf{x}^T(0)P^t \tag{13.2}$$

In particolare, se lo stato al tempo $t = 0$ è S_i e quindi se $\mathbf{x}^T(0) = \mathbf{e}_i^T$, si ha

$$\mathbf{x}^T(t) = \mathbf{e}_i^T P^t,$$

ovvero le probabilità che la catena si trovi in ciascuno degli n stati S_j al tempo t sono le componenti della riga i-esima di P^t.

Osservazione Le probabilità di transizione in 2 passi, da uno stato S_i in uno stato S_k, sono suscettibili di una suggestiva interpretazione sul grafo della catena di Markov. Il passaggio da S_i al tempo $t = 0$ a S_k al tempo $t = 2$ può avvenire secondo n modalità mutuamente esclusive ed esaustive, passando prima dallo stato S_i in $S_j,\ j = 1, 2, \dots, n$, e successivamente dallo stato S_j ad S_k. La j-esima di queste modalità ha probabilità $p_{ij}p_{jk}$, positiva se e solo se entrambi i fattori sono positivi, ovvero se nel grafo c'è un cammino di due archi che connette S_i a S_k passando per S_j. La probabilità complessiva di passare da S_i al tempo

[5] Si sottolinea che il sistema lineare positivo così ottenuto è rappresentato nella forma $\mathbf{x}^T(t+1) = \mathbf{x}^T(t)P$ e che nel suo grafo di influenza un arco dal vertice i al vertice j corrisponde alla presenza di un termine non nullo p_{ij} nella matrice di transizione.

$t = 0$ ad S_k al tempo $t = 2$ è quindi $\sum_j p_{ij}p_{jk}$, elemento di posizione (i, k) nella matrice P^2, corrispondente alla somma di tutte le probabilità associate ai cammini di lunghezza 2 che nel grafo di $\mathcal{C}$ connettono S_i a S_k.
Il ragionamento svolto si estende alla transizione da S_i a S_k in t passi, con $t \geq 3$: sul grafo si dovranno considerare tutti i cammini di lunghezza t che connettono S_i a S_k, cammini corrispondenti agli addendi della sommatoria

$$\sum_{\nu_1,\nu_2,\dots\nu_{t-1}} p_{i,\nu_1}p_{\nu_1,\nu_2}\cdots p_{\nu_{t-1},k}$$

che fornisce $[P^t]_{ik}$, l'elemento in posizione (i, k) nella matrice P^t.

Da quanto precede, appare chiaramente come il comportamento delle catene di Markov sia riconducibile all'analisi delle proprietà delle matrici stocastiche, che costituiscono una classe particolare di matrici positive. L'autovalore massimale λ_0 della matrice di transizione è unitario, essendo unitarie le somme di riga della matrice, e ad esso corrispondono, per le proposizioni 11.5.1 e 11.6.1, uno o più autovettori stocastici sinistri (e uno o più autovettori stocastici destri). Gli autovettori stocastici sinistri costituiscono le *distribuzione stazionarie, o di equilibrio,* della catena: se $\mathbf{p}_0^T$ è uno di essi e se $\mathbf{x}^T(0) = \mathbf{p}_0^T$ è la distribuzione iniziale di $\mathcal{C}$, da $\mathbf{p}_0^T = \mathbf{p}_0^T P$ segue immediatamente $\mathbf{x}^T(t) = \mathbf{p}_0^T$ per ogni $t > 0$. Rimane quindi costante nel tempo la distribuzione di probabilità, e così la probabilità di ciascuno degli stati della catena.
Riassumendo:

Proposizione 13.1.1 [Autovalore massimale e distribuzioni stazionarie] *Se $P \in \mathbb{R}^{n\times n}$ è matrice di transizione della catena $\mathcal{C}$, il suo autovalore massimale è $\lambda_0 = 1$. Ad esso corrisponde almeno un autovettore stocastico sinistro $\mathbf{p}_0^T$, che rappresenta una distribuzione stazionaria di probabilità della catena $\mathcal{C}$, e almeno l'autovettore destro strettamente positivo $\mathbf{1}_n$ (che può essere reso stocastico mediante divisione per n)*

$$\mathbf{p}_0^T P = \mathbf{p}_0^T, \quad P\mathbf{1}_n = \mathbf{1}_n . \qquad \blacksquare$$

13.2 Catene di Markov regolari

Definizione 13.2.1 [Catena regolare] *Una catena di Markov $\mathcal{C}$ si dice "regolare" se la sua matrice di transizione P è primitiva.*

La regolarità, ovvero l'esistenza di un intero positivo h per cui gli elementi $[P^h]_{ik}$ di P^h sono tutti positivi, comporta che qualunque sia lo stato S_i in cui si trova la catena $\mathcal{C}$ all'istante $t = 0$, c'è una probabilità positiva - appunto $[P^h]_{ik}$ - che $\mathcal{C}$ si trovi nello stato S_k al tempo $t = h$, per ogni S_k. Sul grafo della catena ciò equivale ad assumere che per ogni coppia di vertici S_i ed S_k esista almeno un cammino di h archi che connette il primo vertice al secondo (e un cammino di h archi che connette il secondo vertice al primo).

Dal teorema di Perron sulle matrici primitive consegue la seguente

Proposizione 13.2.2 [Comportamento asintotico delle catene regolari] *Sia $\mathcal{C}$ una catena regolare, con matrice di transizione P.*

i) [AUTOVALORE E AUTOVETTORE DOMINANTE] *La matrice P ha l'autovalore dominante $\lambda_0 = 1$ con molteplicità algebrica 1 e tutti gli altri autovalori di P hanno modulo minore di 1. L'autovettore dominante sinistro $\mathbf{p}_0^T$ è strettamente positivo e univocamente determinato dalla condizione di essere stocastico*

$$\begin{aligned} \mathbf{p}_0^T &= [\,\gamma_1 \quad \gamma_2 \quad \cdots \quad \gamma_n\,], \qquad \sum_{i=1}^{n} \gamma_i = \mathbf{p}_0^T \mathbf{1}_n = 1 \\ \mathbf{p}_0^T &= \mathbf{p}_0^T P. \end{aligned}$$

Il vettore $\mathbf{1}_n$ è, a meno di una costante moltiplicativa non nulla, l'unico autovettore dominante destro: $P\mathbf{1}_n = \mathbf{1}_n$.

ii) [RAPPRESENTAZIONE DEI VETTORI STOCASTICI] *Se $\mathbf{p}_0^T, \mathbf{p}_1^T, \ldots, \mathbf{p}_{n-1}^T$ sono una base di Jordan di autovettori e autovettori generalizzati sinistri di P, ogni vettore stocastico $\mathbf{u}^T$ ha proiezione unitaria su $\mathbf{p}_0^T$*

$$\mathbf{u}^T = \mathbf{p}_0^T + \alpha_1 \mathbf{p}_1^T + \ldots + \alpha_{n-1} \mathbf{p}_{n-1}^T, \quad \alpha_i \in \mathbb{C}. \tag{13.3}$$

iii) [DISTRIBUZIONE ASINTOTICA DI PROBABILITÀ] *Se $\mathbf{x}^T(0) = \mathbf{u}^T$ è un'arbitraria distribuzione iniziale di probabilità, la distribuzione $\mathbf{x}^T(t)$ all'istante t soddisfa*

$$\lim_{t\to+\infty} \mathbf{x}^T(t) = \lim_{t\to+\infty} \mathbf{u}^T P^t = \mathbf{p}_0^T \tag{13.4}$$

iv) [MATRICE ASINTOTICA] *Quando t tende a infinito, si ha*

$$\lim_{t\to+\infty} P^t = \begin{bmatrix} \mathbf{p}_0^T \\ \mathbf{p}_0^T \\ \cdots \\ \mathbf{p}_0^T \end{bmatrix} = \mathbf{1}_n \mathbf{p}_0^T := \bar{P} \tag{13.5}$$

La matrice $\bar{P}$
- è idempotente, i.e. $\bar{P}^2 = \bar{P}$,
- commuta con P e la assorbe, i.e. $\bar{P}P = P\bar{P} = \bar{P}$.

PROVA Il punto (i) consegue immediatamente dalla Proposizione 10.4.2: la primitività di P comporta infatti che λ_0 abbia molteplicità algebrica 1 e che gli autovettori dominanti destri e sinistri siano strettamente positivi. Essi sono univocamente individuati quando si imponga loro la condizione di essere stocastici o si specifica il valore di una loro componente.

Per il punto (ii), il teorema di Perron riferito a basi di Jordan sinistre comporta che il vettore stocastico $\mathbf{u}^T = [\,u_1 \quad u_2 \quad \ldots \quad u_n\,]$ sia esprimibile nella forma

$$\mathbf{u}^T = \alpha_0 \mathbf{p}_0^T + \alpha_1 \mathbf{p}_1^T + \ldots + \alpha_{n-1} \mathbf{p}_{n-1}^T, \quad \alpha_0 > 0, \quad \alpha_1, \ldots, \alpha_{n-1} \in \mathbb{C}. \tag{13.6}$$

Postmoltiplicando entrambi i membri di (13.6) per l'autovettore dominante destro $\mathbf{1}_n$ e tenuto conto che, per $i = 1, 2, \ldots, n-1$, $\mathbf{p}_i^T \mathbf{1}_n = 0$, perchè $\mathbf{p}_i^T$ sono autovettori, anche generalizzati, relativi ad autovalori diversi da $\lambda_0 = 1$, si ha

$$1 = \sum_{i=1}^{n} u_i = \mathbf{u}^T \mathbf{1}_n = \alpha_0 \mathbf{p}_0^T \mathbf{1}_n = \alpha_0 \sum_{i=1}^{n} \gamma_i = \alpha_0$$

e quindi (13.3).
Il punto (iii) è una conseguenza del comportamento asintotico di un sistema dotato di autovettore dominante. Poichè gli autovalori diversi dall'autovalore di Perron hanno modulo minore di 1, i trasformati secondo P^t degli autovettori non dominanti, $\mathbf{p}_i^T P^t$, $i = 1, \dots, n-1$, sono infinitesimi al divergere di t e si ha

$$\begin{aligned}
\lim_{t\to+\infty} \mathbf{x}^T(t) &= \lim_{t\to+\infty} \mathbf{u}^T P^t \\
&= \lim_{t\to+\infty} \left(\mathbf{p}_0^T + \alpha_1 \mathbf{p}_1^T + \ldots + \alpha_{n-1}\mathbf{p}_{n-1}^T\right) P^t \\
&= \lim_{t\to+\infty} \left(\mathbf{p}_0^T + \alpha_1 \mathbf{p}_1^T P^t + \ldots + \alpha_{n-1}\mathbf{p}_{n-1}^T P^t\right) \;=\; \mathbf{p}_0^T.
\end{aligned}$$

Per il punto (iv) basta prendere in (13.4) $\mathbf{u}^T = \mathbf{e}_i$, ottenendo $\mathbf{p}_0^T = \lim_{t\to+\infty} \mathbf{e}_i^T P^t$. Pertanto la riga i-esima di P^t converge a $\mathbf{p}_0^T$, e ciò vale per $i = 1, 2, \dots, n$.
Quindi la matrice asintotica $\lim_{t\to+\infty} P^t$ ha tutte le righe eguali a $\mathbf{p}_0^T$ e si esprime nella forma $\mathbf{1}_n \mathbf{p}_0^T$.
L'idempotenza di $\bar{P}$ deriva da

$$\bar{P}^2 = (\mathbf{1}_n \mathbf{p}_0^T)(\mathbf{1}_n \mathbf{p}_0^T) = \mathbf{1}_n(\mathbf{p}_0^T \mathbf{1}_n)\mathbf{p}_0^T = \mathbf{1}_n \mathbf{p}_0^T = \bar{P}$$

mentre da

$$\bar{P} = \lim_{t\to+\infty} P^t = \lim_{t\to+\infty} P^{t+1} = P\left(\lim_{t\to+\infty} P^t\right) = P\bar{P} = \left(\lim_{t\to+\infty} P^t\right) P = \bar{P}P \qquad (13.7)$$

segue che $\bar{P}$ commuta con P e la "assorbe" nel prodotto. ■

Esempio 13.1.1 (Continuazione) La catena di Markov delle condizioni meteorologiche sulla vetta di Utopia è regolare: infatti P^2 è strettamente positiva.
L'autovettore sinistro dominante $\mathbf{p}_0^T$ si ottiene risolvendo in $\gamma_1, \gamma_2, \gamma_3$ l'equazione matriciale

$$[\,\gamma_1 \quad \gamma_2 \quad \gamma_3\,] = [\,\gamma_1 \quad \gamma_2 \quad \gamma_3\,] \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/4 & 1/2 & 1/4 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

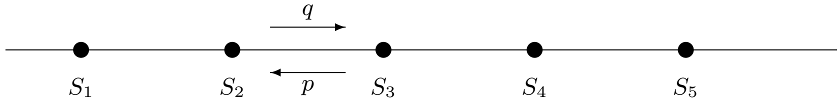
ovvero, eguagliando le prime due componenti di riga (per la terza componente si ottiene un'equazione dipendente dalle prime due), il sistema lineare omogeneo

$$\begin{aligned}
\gamma_1 &= \qquad\;\; \gamma_2/4 + \gamma_3/4 \\
\gamma_2 &= \gamma_1/2 + \gamma_2/2 + \gamma_3/4\,.
\end{aligned}$$

Imponendo che la soluzione sia un vettore stocastico, si ricava

$$\mathbf{p}_0^T = [\,1/5 \quad 2/5 \quad 2/5\,]\,.$$

Esempio 13.2.1 [Passeggiate casuali] Si consideri una particella che si muove fra 5 posizioni S_1, S_2, S_3, S_4, S_5 allineate su un segmento



Si supponga che, se al tempo t la particella si trova in una delle posizioni "interne", all'istante successivo essa si sposti verso la posizione immediatamente a sinistra con probabilità p, $(0 < p < 1)$ o verso la posizione immediatamente a destra con probabilità $q = 1 - p$.

Per quanto attiene al comportamento nelle posizioni estreme, consideriamo varie ipotesi:

1. [ESTREMI ASSORBENTI] la particella vi rimane “intrappolata”;

2. [ESTREMI RIFLETTENTI] se all'istante t la particella si trova in una delle posizioni estreme, con probabilità 1 all'istante $t+1$ viene riflessa nella posizione interna più vicina;

3. [RITORNO AL CENTRO] se all'istante t la particella si trova in una delle posizioni estreme, con probabilità 1 all'istante $t+1$ viene riflessa nella posizione centrale S_3.

Le matrici di transizione nei tre casi sono

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ p & 0 & q & 0 & 0 \\ 0 & p & 0 & q & 0 \\ 0 & 0 & p & 0 & q \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad P_2 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ p & 0 & q & 0 & 0 \\ 0 & p & 0 & q & 0 \\ 0 & 0 & p & 0 & q \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad P_3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ p & 0 & q & 0 & 0 \\ 0 & p & 0 & q & 0 \\ 0 & 0 & p & 0 & q \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

Evidentemente P_1 non è primitiva, e nemmeno irriducibile, poichè in tutte le sue potenze la prima e l'ultima riga rimangono eguali ai vettori $\mathbf{e}_1^T$ ed $\mathbf{e}_5^T$. Non è primitiva nemmeno P_2. Infatti le sue potenze di grado 2,3,4 hanno struttura

$$P_2^2 = \begin{bmatrix} + & 0 & + & 0 & 0 \\ 0 & + & 0 & + & 0 \\ + & 0 & + & 0 & + \\ 0 & + & 0 & + & 0 \\ 0 & 0 & + & 0 & + \end{bmatrix}, \quad P_2^3 = \begin{bmatrix} 0 & + & 0 & + & 0 \\ + & 0 & + & 0 & + \\ 0 & + & 0 & + & 0 \\ + & 0 & + & 0 & + \\ 0 & + & 0 & + & 0 \end{bmatrix}, \quad P_2^4 = \begin{bmatrix} + & 0 & + & 0 & + \\ 0 & + & 0 & + & 0 \\ + & 0 & + & 0 & + \\ 0 & + & 0 & + & 0 \\ + & 0 & + & 0 & + \end{bmatrix},$$

quindi, per $k > 1$, tutte le matrici P_2^{2k+1} e P_2^{2k} hanno rispettivamente la struttura booleana di P_2^3 e di P_2^4. Tuttavia, essendo $I_5 + P_2 + P_2^2 + P_2^3 + P_2^4$ strettamente positiva, la matrice P_2 è irriducibile (quindi, in base alla definizione che daremo in seguito, la catena corrispondente è irriducibile).

P_3 invece è primitiva, risultando $P_3^5 \gg 0$. Per determinarne l'autovettore sinistro dominante $\mathbf{p}_0^T = [\,\gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4 \quad \gamma_5\,]$ si risolve l'equazione

$$[\,\gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4 \quad \gamma_5\,] = [\,\gamma_1 \quad \gamma_2 \quad \gamma_3 \quad \gamma_4 \quad \gamma_5\,] \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ p & 0 & q & 0 & 0 \\ 0 & p & 0 & q & 0 \\ 0 & 0 & p & 0 & q \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

ottenendo le relazioni

$$\begin{aligned} \gamma_1 &= p\gamma_2 = p^2\gamma_3 \\ \gamma_2 &= p\gamma_3 \\ \gamma_4 &= q\gamma_3 \\ \gamma_5 &= q\gamma_4 = q^2\gamma_3 \end{aligned}$$

Il vettore $\mathbf{p}_0^T$ può essere espresso in funzione di γ_3

$$\mathbf{p}_0^T = [\,p^2\gamma_3 \quad p\gamma_3 \quad \gamma_3 \quad q\gamma_3 \quad q^2\gamma_3\,]$$

e successivamente determinato completamente imponendo la condizione di stocasticità

$$(p^2 + p + 1 + q + q^2)\gamma_3 = 1$$

che individua il valore di γ_3

$$\gamma_3 = \frac{1}{1 + (p+q) + (p^2+q^2)} = \frac{1}{3 - 2pq}.$$

- ESERCIZIO 13.2.1 Si ricavino i grafi delle catene associate alle tre diverse situazioni discusse nell'esempio precedente e si verifichino le proprietà di irriducibilità e di primitività.

13.3 Catene di Markov irriducibili

Definizione 13.3.1 [CATENA IRRIDUCIBILE] *Una catena di Markov $\mathcal{C}$ si dice irriducibile (o ergodica) quando la sua matrice di transizione P è irriducibile.*

Poichè le matrici primitive sono irriducibili, le catene regolari costituiscono una sottoclasse della classe delle catene irriducibili. Anche in una catena irriducibile la transizione da S_i a S_k in un opportuno numero di passi ha probabilità positiva, qualunque siano S_i e S_k; tuttavia, a differenza dal caso regolare, non è più garantito che tale probabilità sia positiva per qualsiasi numero di passi, purché sufficientemente grande. Ciò è conseguenza del fatto che, salvo nel caso di una P primitiva, non esiste una potenza m tale che P^m, e con essa $P^{m+1}, P^{m+2} \ldots$, siano tutte strettamente positive. Tuttavia, come vedremo, se η è l'indice di imprimitività di P e se m è abbastanza grande, la somma $P^m + P^{m+1} + \ldots + P^{m+\eta-1}$ risulta strettamente positiva, quindi per tempi sufficientemente grandi ogni stato ha probabilità positiva di essere visitato almeno ogni η passi.

Ricordiamo che, per il corollario 11.5.4, una matrice irriducibile P con indice di imprimitività $\eta > 1$ può essere portata in forma ciclica di Frobenius con una trasformazione di cogredienza. Quindi, eventualmente riordinando gli stati, considereremo come matrice di transizione di una catena irriducibile $\mathcal{C}$ una del tipo

$$P = \begin{bmatrix} 0 & P_{1,2} & 0 & \ldots & 0 \\ 0 & 0 & P_{2,3} & \ldots & 0 \\ & & & \ddots & \\ 0 & 0 & 0 & \ldots & P_{\eta-1,\eta} \\ P_{\eta,1} & 0 & 0 & \ldots & 0 \end{bmatrix}, \tag{13.8}$$

in cui i blocchi diagonali hanno dimensione $\nu_1 \times \nu_1,\ \nu_2 \times \nu_2, \ldots,\ \nu_\eta \times \nu_\eta$. La struttura (non i valori!) delle potenze di P si ripete con periodo η, avendosi per ogni $k \geq 0$,

$$P^{k\eta} = \begin{bmatrix} \star & 0 & 0 & 0 & \ldots & 0 \\ 0 & \star & 0 & 0 & \ldots & 0 \\ 0 & 0 & \star & 0 & \ldots & 0 \\ 0 & 0 & 0 & \star & \ldots & 0 \\ & & & & \ddots & \\ 0 & 0 & 0 & 0 & \ldots & \star \end{bmatrix},\ P^{1+k\eta} = \begin{bmatrix} 0 & \star & 0 & 0 & \ldots & 0 \\ 0 & 0 & \star & 0 & \ldots & 0 \\ 0 & 0 & 0 & \star & \ldots & 0 \\ & & & & \ddots & \\ 0 & 0 & 0 & 0 & \ldots & \star \\ \star & 0 & 0 & 0 & \ldots & 0 \end{bmatrix},\ P^{2+k\eta} = \begin{bmatrix} 0 & 0 & \star & 0 & \ldots & 0 \\ 0 & 0 & 0 & \star & \ldots & 0 \\ & & & & \ddots & \\ 0 & 0 & 0 & 0 & \ldots & \star \\ \star & 0 & 0 & 0 & \ldots & 0 \\ 0 & \star & 0 & 0 & \ldots & 0 \end{bmatrix}, \ldots$$

Se lo stato della catena al tempo $t = 0$ appartiene alla classe $\{S_1, \ldots, S_{\nu_1}\}$, al passo successivo lo stato potrà appartenere solo alla classe $\{S_{\nu_1+1}, \ldots, S_{\nu_1+\nu_2}\}$, nell'istante $t = 2$ solo alla classe $\{S_{\nu_1+\nu_2+1}, \ldots, S_{\nu_1+\nu_2+\nu_3}\}$ e così via. All'istante $t = \eta$ potranno essere rivisitati soltanto stati della prima classe, etc. Le η classi $\{S_1, \ldots, S_{\nu_1}\}$, $\{S_{\nu_1+1}, \ldots, S_{\nu_1+\nu_2}\}$, $\ldots$, $\{S_{\nu_1+\ldots+\nu_{\eta-1}+1}, \ldots, S_n\}$ sono le "classi cicliche" della catena irriducibile $\mathcal{C}$: quando lo stato iniziale appartiene ad una di esse, con il trascorrere del tempo esse vengono visitate ciclicamente. Questa peculiarità non coglie ancora in modo completo il comportamento della catena, perché non specifica secondo quali modalità vengano visitati i singoli stati entro ciascuna classe. Specificamente,

- l'irriducibilità di P garantisce che tutti gli stati della catena abbiano probabilità positiva di essere visitati in qualche istante (ossia: nei blocchi asteriscati delle potenze di P ciascun elemento assume valore positivo per qualche esponente, e anzi per infiniti esponenti);

- da un certo istante in poi tutti gli stati di una classe ciclica hanno simultaneamente probabilità positiva di essere visitati (ovvero nelle potenze di P i blocchi asteriscati diventano strettamente positivi, come conseguenza dalla primitività[6] dei blocchi diagonali di

$$P^\eta = \begin{bmatrix} P_{12}\cdots P_{\eta 1} & & & \\ & P_{23}\cdots P_{12} & & \\ & & \ddots & \\ & & & P_{\eta 1}\cdots P_{(\eta-1)\eta} \end{bmatrix} := \begin{bmatrix} P_{11}^{(\eta)} & & & \\ & P_{22}^{(\eta)} & & \\ & & \ddots & \\ & & & P_{\eta\eta}^{(\eta)} \end{bmatrix})$$

Infatti, per k sufficientemente grande,

a) i blocchi diagonali di $P^{k\eta}$ sono strettamente positivi: l'inizializzazione di $\mathcal{C}$ all'interno di una classe ciclica comporta probabilità positiva di visitare tutti gli stati della medesima classe nell'istante $k\eta$, e nei successivi istanti multipli di η;

b) i blocchi asteriscati in $P^{1+k\eta}, P^{2+k\eta}, \ldots$ sono strettamente positivi (altrimenti P avrebbe una colonna nulla, fatto impossibile per una matrice irriducibile): nell'istante $1 + k\eta$, la catena potrà visitare con probabilità positiva tutti gli stati della classe ciclica successiva a quella in cui $\mathcal{C}$ è stata inizializzata, e nessuno stato delle altre classi, analogamente all'istante $2 + k\eta$ visiterà tutta l'ulteriore classe successiva, etc.

c) se lo stato iniziale S_i appartiene alla prima classe ciclica, per $t = \eta$ la distribuzione di probabilità è fornita dalla i-esima riga di P^η, quindi le sue componenti non nulle appartengono alla riga i-esima del blocco stocastico e primitivo $P_{11}^{(\eta)}$. Una situazione analoga si ripete negli istanti $k\eta$, qualunque sia $k > 0$.
Dalla teoria delle catene regolari sappiamo che

$$\lim_{k\to\infty} \left(P_{11}^{(\eta)}\right)^k = \mathbf{1}_{\nu_1}\mathbf{p}_{01}^T,$$

dove $\mathbf{p}_{01}^T \gg \mathbf{0}^T$ rappresenta la distribuzione asintotica della catena regolare con ν_1 stati e con matrice di transizione $P_{11}^{(\eta)}$. Quindi, negli istanti multipli di η, la distribuzione di probabilità si concentra sugli stati della prima classe ciclica e asintoticamente assume i valori espressi dal vettore strettamente positivo $\mathbf{p}_{01}^T$.

d) Sempre nell'ipotesi che S_i appartenga alla prima classe ciclica, negli istanti $k\eta + 1, k\eta + 2, \ldots$ la distribuzione di probabilità si concentra sugli stati della seconda classe, della terza,... e al divergere di k tende ai vettori strettamente positivi

$$\mathbf{p}_{01}^T P_{12}, \quad \mathbf{p}_{01}^T P_{12}P_{23}, \ \ldots.$$

Si noti che, se P è irriducibile ma non primitiva, l'autovettore stocastico sinistro $\mathbf{p}_0^T$ non rappresenta più la distribuzione asintotica di probabilità della catena per ogni condizione iniziale. In questo caso, il comportamento asintotico di $\mathcal{C}$ dipende dall'insieme di tutti gli autovettori corrispondenti allo spettro periferico e avrà carattere periodico.

Per verificarlo, poniamo $\theta := e^{\frac{j2\pi}{\eta}}$ e siano
$\mathbf{p}_0^T, \mathbf{w}_1^T, \ldots, \mathbf{w}_{\eta-1}^T$ gli autovettori sinistri di P dello spettro periferico $1, \theta, \ldots, \theta^{\eta-1}$
$\mathbf{w}_\eta^T, \ldots, \mathbf{w}_{n-1}^T$ gli altri autovettori (anche generalizzati) sinistri della base di Jordan.

[6] provata nell'esercizio 11.5.2

La distribuzione iniziale di probabilità ha componente unitaria su $\mathbf{p}_0^T$

$$\mathbf{x}^T(0) = \mathbf{p}_0^T + \alpha_1 \mathbf{w}_1^T + \ldots + \alpha_{\eta-1}\mathbf{w}_{\eta-1}^T + \sum_{\nu=\eta}^{n-1} \alpha_\nu \mathbf{w}_\nu^T$$

e al tempo $t \geq 0$ la distribuzione risultante è data da

$$\mathbf{x}^T(t) = \mathbf{x}^T(0)P^t = \mathbf{p}_0^T + \alpha_1 \theta^t \mathbf{w}_1^T + \ldots + \alpha_{\eta-1}\theta^{t(\eta-1)}\mathbf{w}_{\eta-1}^T + \mathbf{h}^T(t), \tag{13.9}$$

in cui, al divergere di t, il termine $\mathbf{h}^T(t)$ tende a zero

$$\lim_{t\to+\infty} \|\mathbf{h}^T(t)\| = 0. \tag{13.10}$$

In (13.9) $\mathbf{h}^T(t)$ è trascurabile rispetto alla somma degli altri termini, dal momento che è unitaria la somma delle componenti di $\mathbf{x}^T(t)$ e da (13.9), (13.10) e da $\theta^\eta = 1$ segue che
– per grandi valori di t si ha $\mathbf{x}^T(t) \sim \mathbf{x}^T(t+\eta)$, ovvero la distribuzione asintotica di probabilità evolve in modo periodico, con periodo[7] η;
– se sommiamo le distribuzioni $\mathbf{x}^T(\cdot)$ su η istanti consecutivi otteniamo[8]

$$\begin{aligned}\sum_{k=0}^{\eta-1}\mathbf{x}^T(t+k) &= \eta\mathbf{p}_0^T + \alpha_1\theta^t\mathbf{w}_1^T\sum_{k=0}^{\eta-1}\theta^k + \alpha_2\theta^{2t}\mathbf{w}_2^T\sum_{k=0}^{\eta-1}\theta^{2k} + \ldots + \alpha_{\eta-1}\theta^{(\eta-1)t}\mathbf{w}_{\eta-1}^T\sum_{k=0}^{\eta-1}\theta^{(\eta-1)k}\\ &+ \sum_{k=0}^{\eta-1}\mathbf{h}^T(t+k) = \eta\mathbf{p}_0^T + \sum_{k=0}^{\eta-1}\mathbf{h}^T(t+k);\end{aligned}$$

quindi, indipendentemente dalla distribuzione iniziale $\mathbf{x}^T(0)$,

$$\lim_{t\to+\infty}\frac{1}{\eta}\sum_{k=0}^{\eta-1}\mathbf{x}^T(t+k) = \mathbf{p}_0^T$$

– scegliendo $\mathbf{x}^T(0) = \mathbf{e}_i,\ i = 1, 2, \ldots, n$, si dimostra, riga per riga, la relazione asintotica

$$\lim_{t\to+\infty}\frac{1}{\eta}\left(P^t + P^{t+1} + \ldots + P^{t+\eta-1}\right) = \mathbf{1}_n\mathbf{p}_0^T \tag{13.11}$$

- Esercizio 13.3.1 Si verifichi che nella catena $\mathcal{C}$ con matrice di transizione

$$P = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1/2 & 0 & 1/2\\ 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

(i) la matrice P (e quindi $\mathcal{C}$) è irriducibile e l'indice di imprimitività è $\eta = 2$;
(ii) se in $t = 0$ lo stato appartiene all'insieme $\{S_1, S_2, S_3\}$ (all'insieme $\{S_4, S_5, S_6\}$) all'istante $t = 8$ è positiva la probabilità di visitare tutti gli stati dell'insieme $\{S_1, S_2, S_3\}$ (dell'insieme $\{S_4, S_5, S_6\}$);
Se in $t = 0$ lo stato iniziale è S_1, in quale stato potrà trovarsi la catena all'istante $t = 1$? quanto tempo si dovrà attendere perche il medesimo stato possa essere rivisitato da $\mathcal{C}$? e perchè possa esserlo per una terza volta? Qual è la distribuzione di probabilità negli istanti $2t$ al divergere di t?

- Esercizio 13.3.2 Si verifichi che, anche nel caso di una catena irriducibile, la matrice $\bar{P} := \mathbf{1}_n\mathbf{p}_0^T$ è idempotente (ossia $\bar{P}^2 = \bar{P}$), commuta con la matrice P e la assorbe (ossia $\bar{P}P = P\bar{P} = \bar{P}$).

[7] Il periodo minimo dipende dalla distribuzione iniziale e può essere η o un suo divisore.
[8] Si noti che, per $\nu = 1, 2, \ldots \eta - 1$, da $0 = (1 - \theta^{\nu\eta}) = (1 - \theta^\nu)\sum_{k=0}^{\eta-1}\theta^{\nu k}$ si ricava $0 = \sum_{k=0}^{\eta-1}\theta^{\nu k}$.

Esempio 13.3.1 [Modello di Ehrenfest] Nel 1907 Paul e Tatiana Ehrenfest introdussero un celebre modello, che porta il loro nome, per spiegare la diffusione nei gas.
(i) Si suppone di avere n particelle identiche e in moto casuale, ripartite fra due urne identiche A e B comunicanti attraverso un'apertura. Il numero delle molecole presenti in una delle urne, p.es. l'urna A, evolve come una catena di Markov, avente per stati gli $n+1$ valori che rappresentano il numero di particelle che possono essere contenute in A e come istanti di transizione gli istanti, indiciati sugli interi, in cui una particella transita attraverso l'apertura. Se all'istante t l'urna A contiene i particelle (e quindi l'urna B ne contiene $n-i$), gli unici stati possibili all'istante $t+1$ sono $i-1$ e $i+1$. Inoltre, la probabilità che una particella si sposti da A a B è i/n, mentre $(n-i)/n$ è la probabilità che una particella si sposti da B ad A. Abbiamo quindi

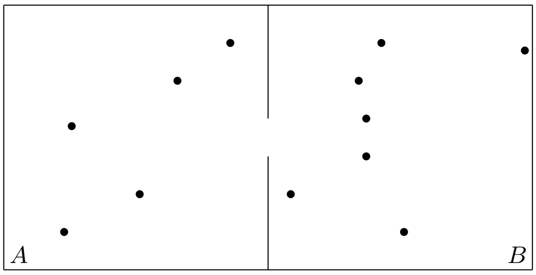


Figura 13.3.1

$$p_{i,i+1} = \frac{n-i}{n}, \qquad p_{i,i-1} = \frac{i}{n}$$

e la matrice di transizione è tridiagonale, con diagonale centrale nulla

$$P = \frac{1}{n}\begin{bmatrix} 0 & n & & & & & & \\ 1 & 0 & n-1 & & & & & \\ & 2 & 0 & n-2 & & & & \\ & & 3 & 0 & n-3 & & & \\ & & & \ddots & \ddots & \ddots & & \\ & & & & & n-1 & 0 & 1 \\ & & & & & & n & 0 \end{bmatrix}. \tag{13.12}$$

(ii) P è irriducibile, perché il grafo corrispondente è fortemente connesso, ma non può essere primitiva: se all'istante $t=0$ lo stato è pari, in tutti gli istanti pari lo stato sarà soltanto pari e negli istanti dispari sarà soltanto dispari. Se la distribuzione iniziale è un vettore canonico, non esiste alcun istante in cui la distribuzione possa essere strettamente positiva e non esiste nessuna potenza strettamente positiva della matrice P. In realtà l'indice di imprimitività della matrice è $\eta = 2$, come si evince immediatamente dal fatto che tutti i circuiti del grafo hanno lunghezza 2. Quindi il secondo autovalore periferico è $\lambda_1 = -1$ ed esiste una matrice diagonale D, a elementi diagonali 1 e -1, per cui risulta (cfr. (11.64)) $P = (-1)D^{-1}PD$ ovvero $DP = -PD$. Una semplice verifica porta a ottenerne la struttura: $D = \operatorname{diag}\{1, -1, 1, \cdots, (-1)^{n-1}, (-1)^n\}$.

(iii) Per determinare l'autovettore sinistro di Perron $\mathbf{p}_0^T$ corrispondente all'autovalore $\lambda_0 = 1$, osserviamo preliminarmente che per $k = 0, 1, \ldots, n$ vale l'identità [9]

$$\frac{n-k}{n}\binom{n}{k} + \frac{k+2}{n}\binom{n}{k+2} = \frac{(n-1)!}{k!(n-k-1)!} + \frac{(n-1)!}{(k+1)!(n-k-2)!} = \frac{n!}{(k+1)!(n-k-1)!} = \binom{n}{k+1}$$

Si verifica allora immediatamente che $\begin{bmatrix} \binom{n}{0} & \binom{n}{1} & \binom{n}{2} & \cdots & \binom{n}{n} \end{bmatrix}$ è l'autovettore sinistro cercato e, normalizzandolo a un vettore stocastico, otteniamo

$$\mathbf{p}_0^T = \frac{1}{2^n}\begin{bmatrix} \binom{n}{0} & \binom{n}{1} & \binom{n}{2} & \cdots & \binom{n}{n} \end{bmatrix}$$

L'autovettore sinistro $\mathbf{w}_1^T$ corrispondente a λ_1 ha elementi di segno alterno, per il corollario 11.5.2

$$\mathbf{w}_1^T = \mathbf{p}_0^T D = \frac{1}{2^n}\begin{bmatrix} \binom{n}{0} & -\binom{n}{1} & \binom{n}{2} & \cdots & (-1)^{n-1}\binom{n}{n-1} & (-1)^n\binom{n}{n} \end{bmatrix},$$

[9] per $i > n$ si pone $\binom{n}{i} = 0$.

come pure ha elementi di segno alterno, ma di eguale modulo, l'autovettore destro $\mathbf{u}^{(1)} = D\,\mathbf{1}_{n+1}$.
(iv) Rappresentiamo un vettore stocastico generico $\mathbf{x}^T(0)$ sulla base di Jordan

$$\mathbf{x}^T(0) = \mathbf{p}_0^T + \alpha_1 \mathbf{w}_1^T + \sum_{i=2}^{n} \alpha_i \mathbf{w}_i^T,$$

in cui $\mathbf{w}_2^T, \mathbf{w}_3^T, \dots, \mathbf{w}_n^T$ sono autovettori[10] relativi ad autovalori a modulo minore di uno. Postmoltiplicando $\mathbf{x}^T(0)$ per $\mathbf{u}^{(1)}$ si ricava che α_1 è la differenza fra la somma delle componenti di indice pari e quella delle componenti di indice dispari di $\mathbf{x}^T(0)$.
In particolare, se $\mathbf{x}^T(0)$ è un vettore canonico, e quindi se all'istante iniziale è certa la ripartizione delle particelle nelle due urne, α_1 varrà $+1$ o -1 a seconda che sia pari o dispari il numero di particelle nell'urna A. Supponendo, p.es., che tale numero sia pari, la distribuzione di probabilità per $t \to +\infty$ sarà data da

$$\begin{aligned}
\mathbf{x}^T(2t) &= \mathbf{p}_0^T + \mathbf{w}_1^T = \frac{1}{2^{n-1}} \begin{bmatrix} \binom{n}{0} & 0 & \binom{n}{2} & 0 & \binom{n}{4} & \dots & \frac{1+(-1)^n}{2}\binom{n}{n} \end{bmatrix} \\
\mathbf{x}^T(2t+1) &= \mathbf{p}_0^T - \mathbf{w}_1^T = \frac{1}{2^{n-1}} \begin{bmatrix} 0 & \binom{n}{1} & 0 & \binom{n}{3} & 0 & \dots & \frac{1-(-1)^n}{2}\binom{n}{n} \end{bmatrix}.
\end{aligned}$$

- Esercizio 13.3.3 Nel modello di Ehrenfest
  (i) se lo stato iniziale è dispari, asintoticamente si trova $\mathbf{x}^T(2t) = \mathbf{p}_0^T - \mathbf{w}_1^T$ e $\mathbf{x}^T(2t+1) = \mathbf{p}_0^T + \mathbf{w}_1^T$;
  (ii) se la distribuzione iniziale ha p come somma delle probabilità degli stati di indice pari, asintoticamente si trova $\mathbf{x}^T(2t) = \mathbf{p}_0^T + (2p-1)\mathbf{w}_1^T$, $\mathbf{x}^T(2t+1) = \mathbf{p}_0^T + (1-2p)\mathbf{w}_1^T$.
- Esercizio 13.3.4* [Analisi modale del modello di Ehrenfest] Tenuto conto della proposizione A.14.4, quali sono i modi del sistema $\mathbf{x}^T(t+1) = \mathbf{x}^T(t)P$, che descrive l'evoluzione della distribuzione di probabilità sugli stati della catena?

13.4 Eventi e variabili casuali in una catena irriducibile

Sia $\mathcal{C}$ una catena irriducibile, con n stati $S_1, S_2, \dots, S_n$ e matrice di transizione $P \in \mathbb{R}_+^{n\times n}$.

13.4.1 Traiettorie

Se la distribuzione di probabilità iniziale è $\mathbf{x}_0^T = [\,\xi_1 \quad \xi_2 \quad \dots \quad \xi_n\,]$, la probabilità che nelle prime t transizioni la catena segua una specifica traiettoria

$$S(0) = S_{i_0},\ S(1) = S_{i_1},\ \dots, S(t) = S_{i_t}$$

è data, per la proprietà di markovianità, da

$$\begin{aligned}
&\mathrm{pr}\{S(0) = S_{i_0}\}\,\mathrm{pr}\{S(1) = S_{i_1} | S(0) = S_{i_0}\}\,\mathrm{pr}\{S(2) = S_{i_2} | S(1) = S_{i_1}\}\cdots \\
&\cdots \mathrm{pr}\{S(t) = S_{i_t} | S(t-1) = S_{i_{t-1}}\} = \xi_{i_0}\, p_{i_0,i_1}\, p_{i_1,i_2} \cdots p_{i_{t-1},i_t} \qquad (13.13)
\end{aligned}$$

Come caso particolare, si ottiene la probabilità che la catena rimanga nello stato S_{i_0} fino all'istante t

$$\mathrm{pr}\{S(0) = S(1) = \dots = S(t) = S_{i_0}\} = \xi_{i_0} p_{i_0,i_0}^t \qquad (13.14)$$

e, essendo $p_{i_0,i_0} < 1$ per l'irriducibilità di P, al divergere di t la probabilità di rimanere costantemente nello stato S_{i_0} tende a zero.

[10] La matrice P ha $n+1$ autovalori distinti (si veda la matrice di Marc Kac in A.14), quindi non ci sono autovettori generalizzati.

La (13.14) si generalizza. Sia A_1 un sottoinsieme di cardinalità $r < n$ dell'insieme degli stati, sottoinsieme che, previo un riordino degli stati, non è restrittivo supporre composto da $S_1, S_2, \ldots, S_r$. Per studiare la permanenza della catena in A_1, conviene partizionare conformemente la matrice di transizione P e le sue potenze P^t

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}, \qquad P^t = \begin{bmatrix} P_{11}^{(t)} & P_{12}^{(t)} \\ P_{21}^{(t)} & P_{22}^{(t)} \end{bmatrix} \tag{13.15}$$

e premettere alcune considerazioni sul blocco $P_{11} \in \mathbb{R}_+^{r\times r}$.
Esso ha raggio spettrale minore di 1, per il corollario 11.6.2, quindi P_{11}^t è infinitesimo per $t \to +\infty$. Inoltre, per i compreso fra 1 e r, la somma dei termini della riga i-esima di P_{11}^t

- è decrescente[11] al crescere di t, essendo $\mathbf{e}_i^T P_{11}^{t+1} \mathbf{1}_r = \mathbf{e}_i^T P_{11}^t (P_{11} \mathbf{1}_r) \leq \mathbf{e}_i^T P_{11}^t \mathbf{1}_r$,
- è minore di 1 per qualche $t < n$: infatti, per ogni j compreso fra $r + 1$ e n, esiste (per il punto 5 della proposizione 11.2.2) un esponente $t < n$ tale che $[P^t]_{ij} > 0$, e ciò comporta
$$\mathbf{e}_i^T P_{11}^t \mathbf{1}_r \leq \mathbf{e}_i^T P_{11}^{(t)} \mathbf{1}_r < 1$$

Quindi in P_{11}^{n-1} le somme di riga sono tutte minori di 1

$$1 - \gamma := \max_{i=1,2,\ldots,r} \mathbf{e}_i^T P_{11}^{n-1} \mathbf{1}_r < 1 \tag{13.16}$$

Le probabilità che all'istante 0 la catena si trovi nei vari stati del sottoinsieme A_1 sono le componenti del vettore

$$[\, \xi_1 \quad \xi_2 \quad \ldots \quad \xi_r \,]. \tag{13.17}$$

Vogliamo verificare che le componenti del vettore

$$[\, \eta_1 \quad \eta_2 \quad \ldots \quad \eta_r \,] = [\, \xi_1 \quad \xi_2 \quad \ldots \quad \xi_r \,] P_{11}^t \tag{13.18}$$

rappresentano le probabilità che $\mathcal{C}$ appartenga al sottoinsieme A_1 negli istanti $0, 1, \ldots, t-1$ e che si trovi in uno degli stati $S_1, S_2, \ldots, S_r$ al tempo t. Se assumiamo induttivamente vera la (13.18), l'eventualità che sia $S(t+1) = S_j \in A_1$, con $S(\sigma)$ in A_1 per $\sigma = 0, 1, \ldots, t$, può verificarsi secondo r modalità, mutuamente esclusive ed esaustive di tutte le possibilità:
- $\mathcal{C}$ si trova in $S(t) = S_1$, con $S(\sigma) \in A_1, \forall \sigma < t$, e transita in S_j nell'istante $t + 1$,
- $\mathcal{C}$ si trova in $S(t) = S_2$, con $S(\sigma) \in A_1, \forall \sigma < t$, e transita in S_j nell'istante $t + 1$,
- $\ldots$
- $\mathcal{C}$ si trova in $S(t) = S_r$, con $S(\sigma) \in A_1, \forall \sigma < t$, e transita in S_j nell'istante $t + 1$

e la probabilità complessiva è $\sum_{i=1}^r \eta_i p_{i,j}$, ovvero è la componente j-esima del vettore $[\, \eta_1 \quad \eta_2 \quad \ldots \quad \eta_r \,] P_{11} = [\, \xi_1 \quad \xi_2 \quad \ldots \quad \xi_r \,] P_{11}^{t+1}$. Quindi (13.18) è corretta.

Proposizione 13.4.1 [Permanenza in un insieme] *Sia $\mathcal{C}$ una catena irriducibile, con n stati $S_1, S_2, \ldots, S_n$, distribuzione iniziale $\mathbf{x}_0^T = [\, \xi_1 \quad \xi_2 \quad \ldots \quad \xi_n \,]$ e matrice di transizione P partizionata come in (13.15).*

i) La probabilità che in ogni istante dell'intervallo $[0, t]$ lo stato della catena appartenga al sottoinsieme $A_1 = \{S_1, S_2, \ldots, S_r\}$ con $0 < r < n$, è data da

$$[\, \xi_1 \quad \xi_2 \quad \ldots \quad \xi_r \,] P_{11}^t \mathbf{1}_r \ . \tag{13.19}$$

[11] in senso lato

ii) Esiste $\gamma > 0$ tale che, per ogni distribuzione di probabilità all'istante $t = 0$, la probabilità che $\mathcal{C}$ si trovi in A_1 durante tutto l'intervallo $[1, n]$ non eccede $1 - \gamma$.

iii) La probabilità che $\mathcal{C}$ si trovi in A_1 durante tutto l'intervallo $[1, t]$ tende a zero al divergere di t.

Prova (i) è conseguenza immediata di (13.18).
ii) Qualunque sia la distribuzione di probabilità $\mathbf{x}^T(0) = [\,\mathbf{x}_1^T \quad \mathbf{x}_2^T\,]$, con $\mathbf{x}_1 \in \mathbb{R}^r$, si ha $\mathbf{x}^T(1) = [\,\mathbf{x}_1^T P_{11} + \mathbf{x}_2^T P_{21} \quad \mathbf{x}_1^T P_{12} + \mathbf{x}_2^T P_{22}\,]$ e il blocco

$$[\,\xi_1' \quad \xi_2' \quad \dots \quad \xi_r'\,] = \mathbf{x}_1^T P_{11} + \mathbf{x}_2^T P_{21}$$

fornisce le probabilità che al tempo $t = 1$ la catena si trovi nei diversi stati di A_1.
Applicando il ragionamento che ha portato a (13.18), si vede che ciascuna componente della riga $[\,\xi_1' \quad \xi_2' \quad \dots \quad \xi_r'\,]\, P_{11}^{t-1}$ fornisce la probabilità che la catena, dopo aver visitato soltanto stati di A_1 nell'intervallo $[1, t-1]$, si trovi al tempo t in uno specifico stato di A_1, quindi

$$[\,\xi_1' \quad \xi_2' \quad \dots \quad \xi_r'\,]\, P_{11}^{n-1}\mathbf{1}_r \;=\; \sum_{i=1}^{r} \xi_i'\, \mathbf{e}_i^T P_{11}^{n-1}\mathbf{1}_r \;\le\; \sum_{i=1}^{r} \xi_i' \Big(\max_{i=1,2,\dots,r} \mathbf{e}_i^T P_{11}^{n-1}\mathbf{1}_r \Big) \;\le\; 1-\gamma \;<\; 1$$

fornisce la probabilità complessiva che la catena nell'intervallo $[1, n]$ abbia visitato soltanto stati di A_1.
iii) P_{11}^t è infinitesima al divergere di t , perciò la probabilità $[\,\xi_1' \quad \xi_2' \quad \dots \quad \xi_r'\,]\, P_{11}^{n-1}\mathbf{1}_r$ che la catena visiti in $[1, t]$ soltanto gli stati di A_1 è anch'essa infinitesima. ■

L'evento che $\mathcal{C}$ rimanga nell'insieme A_1 in ogni istante $t > 0$ e, *a fortiori*, in ogni istante $t \ge 0$ ha probabilità nulla. Equivalentemente, ha probabilità nulla l'evento che la catena non "visiti" mai, per nessun $t > 0$, l'insieme complementare $A_2 = \{S_{r+1}, S_{r+2}, \dots, S_n\}$. In particolare ha probabilità nulla, indipendentemente dallo stato (o dalla distribuzione) iniziale, l'evento che una catena irriducibile non visiti per nessun $t > 0$ uno specifico stato S_k (basta scegliere $A_2 = \{S_k\}$).

- Esercizio 13.4.1 Con riferimento alla partizione degli stati nei sottoinsiemi $A_1 = \{S_1, \dots, S_r\}$ e $A_2 = \{S_{r+1}, \dots, S_n\}$, entrambi non vuoti, si consideri la matrice partizionata (13.15) e si ponga $\mathbf{x}^T(0) = [\,\mathbf{x}_1^T \quad \mathbf{x}_2^T\,]$.
(i) La probabilità che la catena si trovi in A_1 negli istanti 0,1,4,5 e in A_2 negli istanti 2 e 3 è data da $\mathbf{x}_1^T P_{11} P_{12} P_{22} P_{21} P_{11} \mathbf{1}_r$.
(ii) Si generalizzi il risultato ad arbitrarie successioni finite $S(0) \in A_{i_0},\ S(1) \in A_{i_1},\ \dots,\ S(N) \in A_{i_t}$.
(iii) Se $\mathcal{C}$ è irriducibile, non è necessariamente vero che, data una successione (infinita) di vincoli sulle traiettorie $S(0) \in A_{i_0},\ S(1) \in A_{i_1},\ S(2) \in A_{i_2} \dots \ A_{i_\nu} \in \{A_1, A_2\}$, si abbia

$$\mathrm{pr}\,\{S(0) \in A_{i_0},\ S(1) \in A_{i_1},\ S(2) \in A_{i_2},\ \dots\dots\} = 0.$$

E se la catena $\mathcal{C}$ è regolare?

♯ *Soluzione per (iii) : si considerino la matrice irriducibile* $P = \begin{bmatrix} 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{bmatrix}$, *la distribuzione iniziale* $\mathbf{x}(0)^T = [\,1/4 \quad 1/4 \quad 1/4 \quad 1/4\,]$ *e l'insieme* $A_1 = \{1, 2\}$. *La probabilità di avere* $S(2k) \in A_1, S(2k+1) \in A_2,\ \forall k \le N$, *non converge a zero al crescere di* N.

Se $\mathcal{C}$ *è regolare, i.e. con* P *primitiva, esiste* h *tale da aversi* $M := P^h = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \gg 0$. *Allora ogni blocco* $M_{\mu,\nu},\ \ \mu, \nu = 1, 2$ *è esprimibile nella forma*

$$M_{\mu,\nu} = \sum_{i_1,i_2,\ldots,i_{h-1}\in\{1,2\}} P_{\mu,i_1} P_{i_1,i_2} \ldots P_{i_{h-1},\nu}$$

Le righe di ciascuno dei blocchi $M_{\mu,\nu}$ hanno tutte somma degli elementi minore di 1, quindi per un opportuno $\epsilon \in (0,1)$ anche minore di $1-\epsilon$. A maggior ragione, in ciascuno dei prodotti $P_{\mu,i_1} P_{i_1,i_2} \ldots P_{i_{h-1},\nu}$ le righe hanno tutte somma degli elementi minore di $1-\epsilon$.

Poiché i quattro blocchi di P hanno somme di riga non superiori a 1, un prodotto di t blocchi di P, con $t = qh + r$, $0 \le r < h$ ha somme di riga minori di $(1-\epsilon)^q$ per l'esercizio 11.6.5. Al divergere di t ogni prodotto di t blocchi di P converge a zero, quindi tende a zero $\mathrm{pr}\,\{S(0) \in A_{i_0},\ S(1) \in A_{i_1},\ S(2) \in A_{i_2},\ \ldots\ldots\} = 0$ *, comunque si scelga la successione di vincoli sulle traiettorie.*

13.4.2 Reversibilità

Supponiamo che $\mathbf{p}_0^T = [\,\gamma_1 \quad \gamma_2 \quad \ldots \quad \gamma_n\,] \gg \mathbf{0}^T$ sia la distribuzione di equilibrio della catena irriducibile $\mathcal{C}$.

Se $\mathcal{C}$ è inizializzata all'equilibrio, ovvero se $\mathbf{x}^T(0) = \mathbf{p}_0^T$, l'evento costituito dal passaggio della catena, nell'intervallo $[0, N]$, attraverso uno specificato insieme di stati $S(0) = S_{i_0},\ S(1) = i_1,\ S(2) = i_2, \ldots,\ S(N) = i_N$ ha probabilità (cfr. 13.13)

$$\gamma_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{N-1} i_N},$$

ovvero è il prodotto dei pesi degli archi del cammino che porta da $S(0)$ a $S(N)$, attraverso gli stati che abbiamo scelto, e della probabilità γ_{i_0} che all'istante iniziale lo stato sia S_{i_0}.

Diremo che una catena $\hat{\mathcal{C}}$ con n stati $\{S_1, S_2, \ldots, S_n\}$ e matrice di transizione $\hat{P}$ realizza l'inversione temporale della catena $\mathcal{C}$ (o è la *catena inversa* di $\mathcal{C}$) se, scelto un arbitrario $N > 0$, ogni cammino di lunghezza N da $S(0)$ a $S(N)$ nella catena $\mathcal{C}$ inizializzata dalla distribuzione $\mathbf{p}_0^T$ ha la medesima probabilità del cammino "inverso" nella catena $\hat{\mathcal{C}}$ inizializzata da $\mathbf{p}_0^T$:

$$\begin{aligned} \mathrm{pr}\{S(0) = S_{i_0}, S(1) = S_{i_1}, \ldots, S(N) = S_{i_N}\} &= \gamma_{i_0} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{N-1} i_N} \\ = \mathrm{pr}\{\hat{S}(0) = S_{i_N}, \hat{S}(1) = Si_{N-1}, \ldots, \hat{S}(N) = S_{i_0}\} &= \gamma_{i_N} \hat{p}_{i_N i_{N-1}} \hat{p}_{i_{N-1} i_{N-2}} \cdots \hat{p}_{i_1 i_0} \end{aligned} \tag{13.20}$$

Proposizione 13.4.2 [Esistenza e proprietà della catena inversa] *Sia $\mathcal{C}$ una catena irriducibile con matrice di transizione P e distribuzione di equilibrio $\mathbf{p}_0^T \gg \mathbf{0}^T$. Posto $\mathbf{p}_0^T = [\,\gamma_1 \quad \ldots \quad \gamma_n\,]$, $D = \mathrm{diag}\{\gamma_1, \ldots, \gamma_n\}$ e $Q = DPD^{-1}$,*

i) la matrice $\hat{P} := Q^T$ è stocastica irriducibile, con distribuzione di equilibrio $\mathbf{p}_0^T$;

ii) la catena $\hat{\mathcal{C}}$ associata a $\hat{P}$ realizza l'inversione temporale di $\mathcal{C}$ ed è l'unica che la realizza.

Prova (i) Evidentemente Q è una matrice positiva ed ha il medesimo grafo di P. Quindi Q e Q^T sono irriducibili. Per verificare che $\hat{P} = Q^T$ è una matrice stocastica, si noti che

$$\hat{P}\mathbf{1}_n = Q^T \mathbf{1}_n = D^{-1} P^T D \mathbf{1}_n = D^{-1} P^T \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix} = D^{-1} \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{bmatrix} = \mathbf{1}_n \ ,$$

mentre
$$\mathbf{p}_0^T \hat{P} = \mathbf{p}_0^T D^{-1} P^T D = \mathbf{1}_n^T P^T D = \mathbf{1}_n^T D = \mathbf{p}_0^T \tag{13.21}$$
dimostra che $\mathbf{p}_0^T$ è la distribuzione di equilibrio di $\hat{P}$.
(ii) La relazione $\hat{P}^T D = DP$ equivale alle condizioni scalari

$$\gamma_j \hat{p}_{ji} = \gamma_i p_{ij}, \ \forall i, j \tag{13.22}$$

che, sostituite nel secondo membro delle (13.20),

$$\begin{aligned}
\gamma_{i_N} \hat{p}_{i_N i_{N-1}} \hat{p}_{i_{N-1} i_{N-2}} \cdots \hat{p}_{i_2 i_1} \hat{p}_{i_1 i_0} &= \gamma_{i_N} \hat{p}_{i_N i_{N-1}} \hat{p}_{i_{N-1} i_{N-2}} \cdots \ \hat{p}_{i_3 i_2} \hat{p}_{i_2 i_1} \frac{\gamma_{i_0}}{\gamma_{i_1}} p_{i_0 i_1} \\
&= \gamma_{i_N} \hat{p}_{i_N i_{N-1}} \hat{p}_{i_{N-1} i_{N-2}} \cdots \ \hat{p}_{i_3 i_2} \frac{\gamma_{i_0}}{\gamma_{i_2}} p_{i_0 i_1} p_{i_1 i_2} \\
&= \qquad \cdots \quad \cdots \\
&= \gamma_{i_N} \frac{\gamma_{i_0}}{\gamma_{i_N}} p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{N-2} i_{N-1}} p_{i_{N-1} i_N}
\end{aligned}$$

forniscono il primo. Quindi $\hat{P} = \left(DPD^{-1}\right)^T$ è matrice di transizione di una catena che realizza l'inversione temporale di $\mathcal{C}$. Per verificare l'unicità di tale catena, basta osservare che le condizioni (13.20), ristrette a tutti i cammini di lunghezza unitaria, sono proprio le condizioni $\gamma_i p_{ij} = \gamma_j \hat{p}_{ji}, \ \ \forall i, j$, che equivalgono, come s'è detto, a $\hat{P}^T = DPD^{-1}$. ■

Definizione 13.4.3 [Catene reversibili] *Una catena irriducibile $\mathcal{C}$ con matrice di transizione P e distribuzione di equilibrio $\mathbf{p}_0^T = [\,\gamma_1 \quad \gamma_2 \quad \ldots \quad \gamma_n\,]$ si dice reversibile se coincide con la sua catena inversa $\hat{\mathcal{C}}$, ovvero se la matrice P soddisfa la condizione*

$$P^T D = DP, \qquad D := \operatorname{diag}\{\gamma_1, \gamma_2, \ldots, \gamma_n\}. \tag{13.23}$$

Esempio 13.4.1 [Conduttanze di un grafo privo di autoanelli e passeggiata casuale]
Si consideri un grafo non orientato $\mathcal{G} = (\mathcal{V}, \mathcal{L})$. $\mathcal{V} = \{S_1, S_2, \ldots, S_n\}$ o, per brevità, $\{1, 2, \ldots, n\}$, è l'insieme dei vertici e $\mathcal{L}$ è l'insieme dei lati, che supponiamo privo di autoanelli. Ogni lato (non orientato!) può essere identificato con il sottoinsieme $\{S_i, S_j\}$ di $\mathcal{V}$ costituito dai due vertici che esso connette.
Si suppone che il grafo sia *connesso*, ovvero che due vertici qualsiasi siano connessi da un cammino (non orientato!). Ogni lato $\{S_i, S_j\}$ è caratterizzato da una costante positiva, che indicheremo con g_{ij} (o con g_{ji}) e chiameremo *conduttanza* del lato, interpretandola, a livello intuitivo, come una misura della pervietà al flusso, in entrambe le direzioni.
Se trasformiamo ciascun lato $\{S_i, S_j\}$ nella coppia di archi orientati (S_i, S_j) e (S_j, S_i), otteniamo da $\mathcal{G}$ un grafo orientato $\mathcal{D}$ "bidiretto" (nel senso che se si va da S_i a S_j in un passo, in un passo si va anche da S_j a S_i) ed etichettato in modo simmetrico (le etichette g_{ij} di (S_i, S_j) e g_{ji} di (S_j, S_i) coincidono entrambe con la conduttanza del lato $\{S_i, S_j\}$: si veda la figura 13.4.1).
Chiamiamo *intorno del vertice* S_i il sottoinsieme $\mathcal{N}_i$, costituito dai vertici S_k che in $\mathcal{D}$ sono connessi al vertice S_i da un arco (S_i, S_k) e associamo a $\mathcal{D}$ una catena di Markov, definendo le probabilità di transizione mediante le conduttanze degli archi:
$$p_{ij} = \frac{g_{ij}}{\sum_{k \in \mathcal{N}_i} g_{ik}} \tag{13.24}$$

La probabilità di passare dallo stato S_i allo stato S_j è il rapporto fra la conduttanza dell'arco orientato (S_i, S_j) e la somma di tutte e conduttanze degli archi che connettono S_i ai vertici del suo intorno: giunti al vertice S_i in una passeggiata casuale lungo gli archi del grafo, la probabilità di percorrere una arco piuttosto che un altro, fra quelli che si dipartono da S_i, è proporzionale alle

rispettive conduttanze.
La matrice $P = [p_{ij}]$ in generale non è simmetrica, mentre lo è la matrice $G := [g_{ij}]$, ma, soddisfacendo

$$P = \mathrm{diag}\{(\sum_{j=1}^{n} g_{1j})^{-1}, (\sum_{j=1}^{n} g_{2j})^{-1}, \ldots, (\sum_{j=1}^{n} g_{nj})^{-1}\}G,$$

è stocastica per righe, ed è irriducibile perché è associata ad un grafo orientato connesso. Per ottenere la distribuzione stazionaria, si ricava un autovettore di Perron di P

$$\begin{aligned}
&\left[\, \sum_{j=1}^{n} g_{1j} \quad \sum_{j=1}^{n} g_{2j} \quad \cdots \quad \sum_{j=1}^{n} g_{nj} \,\right] P \\
= \;& \left[\, \sum_{j=1}^{n} g_{1j} \quad \sum_{j=1}^{n} g_{2j} \quad \cdots \quad \sum_{j=1}^{n} g_{nj} \,\right] \mathrm{diag}\Big\{(\sum_{j=1}^{n} g_{1j})^{-1}, (\sum_{j=1}^{n} g_{2j})^{-1}, \ldots, (\sum_{j=1}^{n} g_{nj})^{-1}\Big\}G \\
= \;& \mathbf{1}^T G = \mathbf{1}^T G^T = \left[\, \sum_{j=1}^{n} g_{1j} \quad \sum_{j=1}^{n} g_{2j} \quad \cdots \quad \sum_{j=1}^{n} g_{nj} \,\right]
\end{aligned}$$

e lo si normalizza:

$$\mathbf{p}_0^T = [\, \gamma_1 \quad \gamma_2 \quad \ldots \quad \gamma_n \,] = \frac{1}{\sum_{i,j=1}^{n} g_{ij}} \left[\, \sum_{j=1}^{n} g_{1j} \quad \sum_{j=1}^{n} g_{2j} \quad \cdots \quad \sum_{j=1}^{n} g_{nj} \,\right] \tag{13.25}$$

Vogliamo verificare infine che la passeggiata casuale sul grafo è una catena reversibile, ovvero vale (13.23). L'identità $P^T D = DP$ equivale infatti, per ogni i e k tali che $g_{ik} \neq 0$, alle identità scalari

$$\begin{aligned}
&[P^T]_{ik}\gamma_k = \gamma_i[P]_{ik} \quad \Leftrightarrow \quad [P]_{ki}\gamma_k = \gamma_i[P]_{ik} \quad \Leftrightarrow \quad g_{ki}(\sum_j g_{kj})^{-1}\gamma_k = \gamma_i g_{ik}(\sum_j g_{ij})^{-1} \\
&\Leftrightarrow \quad \frac{\gamma_k}{\gamma_i} = \frac{\sum_j g_{kj}}{\sum_j g_{ij}}
\end{aligned}$$

e le eguaglianze delle seconda riga sono verificate grazie alla struttura di $\mathbf{p}_0^T$.

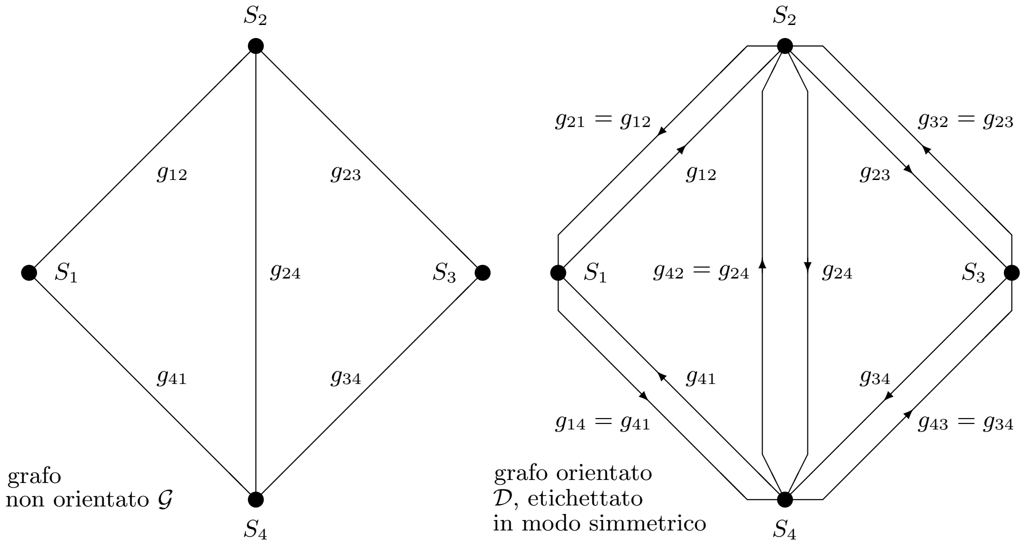


Figura 13.4.1

- Esercizio 13.4.2 Se la catena di Markov della passeggiata casuale dell'esempio 13.4.1 non è regolare, allora P è irriducibile con indice di imprimitività 2.
- Esercizio 13.4.3 Sia $\mathcal{C}$ una catena di Markov irriducibile, reversibile, con distribuzione stazionaria $\mathbf{p}_0^T = [\, \gamma_1 \quad \gamma_2 \quad \ldots \quad \gamma_n \,]$, e si supponga che la matrice di transizione $P = [p_{ij}]$ abbia diagonale nulla. Allora $\mathcal{C}$ è una passeggiata casuale su un grafo privo di autoanelli nel quale le conduttanze dei lati sono definite da $g_{ij} = \gamma_i p_{ij}$.

13.4.3 Tempi di attesa

Consideriamo la variabile casuale a valori interi[12] f_{ik}, che fornisce il primo istante successivo a $t = 0$ in cui la catena $\mathcal{C}$, che parte dallo stato S_i all'istante $t = 0$, raggiunge lo stato S_k. Il resto del paragrafo è dedicato all'analisi di alcuni caratteri statistici di f_{ik}.

Proposizione 13.4.4 [f_{ik} HA MEDIA FINITA] *Si consideri una catena irriducibile $\mathcal{C}$ e si indichi con m_{ik} il valor medio della variabile casuale f_{ik} che fornisce il primo istante, successivo all'istante iniziale, in cui la catena, partendo in S_i, entra nello stato S_k:*

$$\mathbf{E}(f_{ik}) := m_{ik} = \sum_{\nu=1}^{+\infty} \nu \; \mathrm{pr}\{f_{ik} = \nu\} \; . \tag{13.26}$$

Si ha sempre $m_{ik} < \infty$.

PROVA Trattandosi di una serie a termini non negativi, la somma della serie (13.26) può essere valutata accorpando i termini a gruppi di n

$$m_{ik} = \sum_{\nu=1}^{n} \nu \; \mathrm{pr}\{f_{ik} = \nu\} + \sum_{\nu=n+1}^{2n} \nu \; \mathrm{pr}\{f_{ik} = \nu\} + \sum_{\nu=2n+1}^{3n} \nu \; \mathrm{pr}\{f_{ik} = \nu\} + \ldots$$

Per maggiorare la seconda sommatoria, osserviamo che il primo passaggio per S_k potrà verificarsi in un istante ν compreso fra $n+1$ e $2n$ solo se in nessun istante compreso fra 1 e n la catena avrà visitato lo stato S_k. Quindi la probabilità che il primo passaggio si verifichi nell'istante ν compreso fra $n+1$ e $2n$ è non superiore a quella che nell'intervallo $[1, n]$ lo stato S_k non sia mai stato visitato. Tale evento, qualunque sia lo stato iniziale della catena, ha probabilità non superiore a $1 - \gamma$, per un opportuno $\gamma > 0$ (si veda la proposizione 13.4.1). Con un ragionamento analogo, nella terza sommatoria le probabilità che il primo passaggio per S_k avvenga in un istante ν compreso fra $2n+1$ e $3n$ non supera $(1-\gamma)^2$, etc. Si ha quindi

$$\begin{aligned} m_{ik} \;&\leq\; \sum_{\nu=1}^{n} \nu + \sum_{\nu=n+1}^{2n} \nu(1-\gamma) + \sum_{\nu=2n+1}^{3n} \nu((1-\gamma)^2 + \ldots \\ &\leq\; n^2 + 2n^2(1-\gamma) + 3n^2(1-\gamma)^2 + \ldots \;=\; n^2 \Big(1 + 2(1-\gamma) + 3(1-\gamma)^2 + \ldots\Big) \end{aligned}$$

in cui la serie dell'ultima riga converge per il criterio del rapporto. ■

Vogliamo ora determinare esplicitamente, per ogni coppia (i, k), il valore (finito) di m_{ik}. A tale scopo, riunite tutte le medie in un'unica matrice positiva $M = [m_{ik}]$ di dimensione $n \times n$, procederemo secondo i passi seguenti:

1. ricaveremo un'equazione lineare matriciale che deve essere soddisfatta dalla matrice delle medie M;
2. dimostreremo che tutte le matrici che risolvono l'equazione hanno la medesima diagonale, che ricaveremo esplicitamente;
3. verificheremo che è unica la soluzione dell'equazione del punto 1;

[12] Per $i \neq k$ la variabile f_{ik} è l'istante di "primo passaggio" attraverso lo stato S_k; per $i = k$ l'istante di primo passaggio è l'istante $t = 0$, mentre f_{ii} fornisce il "tempo di ricorrenza", ovvero il primo istante in cui la catena ritorna nello stato S_i da cui è partita.

4. negli esercizi, si descriverà la struttura esplicita della matrice delle medie in funzione dalla matrice di transizione P e dalla distribuzione $\mathbf{p}_0^T$.

Proposizione 13.4.5 [Equazione per la matrice delle medie] *Si consideri una catena irriducibile con matrice di transizione P e sia $M = [m_{ik}]$ la matrice dei valori medi delle variabili casuali f_{ik}.*
Allora M è soluzione della seguente equazione nella matrice incognita X,

$$X = \mathbf{1}_n \mathbf{1}_n^T + P\left(X - \mathrm{diag}(X)\right). \tag{13.27}$$

dove X e hanno dimensione $n \times n$, e $\mathrm{diag}(X) = \mathrm{diag}\{[X]_{11}, [X]_{22}, \dots, [X]_{nn}\}$.

Prova Se lo stato iniziale della catena è $S(0) = S_i$, la media m_{ik} della variabile f_{ik} si può ottenere come media delle sue medie condizionate dall'esito del primo passo della catena. In altri termini, si assumerà la condizione che al primo passo la catena $\mathcal{C}$ visiti lo stato $S_j,\ j = 1, 2, \dots, n$ e si peseranno con le probabilità p_{ij} (di transizione in un passo) i tempi medi per la prima visita dello stato S_k sotto la condizione $S(1) = S_j$:

$$\begin{aligned} m_{ik} &= \sum_{j=1}^{n} \mathbf{E}(f_{ik}|S(1) = S_j)\mathrm{pr}\{S(1) = S_j\} \\ &= \mathbf{E}(f_{ik}|S(1) = S_k) \cdot \mathrm{pr}\{S(1) = S_k\} + \sum_{j \neq k} \mathbf{E}(f_{ik}|S(1) = S_j) \cdot \mathrm{pr}\{S(1) = S_j\} \\ &= 1 \cdot p_{ik} + \sum_{j \neq k} (1 + m_{jk}) p_{ij}. \end{aligned} \tag{13.28}$$

Ma allora per ogni i e k valgono le uguaglianze

$$m_{ik} = \sum_{j=1}^{n} p_{ij} + \sum_{j \neq k} p_{ij} m_{jk} = 1 + \sum_{j=1}^{n} p_{ij} m_{jk} - p_{ik} m_{kk} = 1 + [PM]_{ik} - [P\mathrm{diag}(M)]_{ik}$$

che in termini matriciali si compendiano nella scrittura

$$M = \mathbf{1}_n \mathbf{1}_n^T + P\left(M - \mathrm{diag}(M)\right). \tag{13.29}$$

Quindi M è soluzione dell'equazione (13.27). ■

Lemma 13.4.6 [Unicità della diagonale nelle soluzioni dell'equazione] *Sia $\mathbf{p}_0^T = [\,\gamma_1 \quad \gamma_2 \quad \dots \quad \gamma_n\,]$ la distribuzione stazionaria della catena irriducibile $\mathcal{C}$ e sia $\tilde{M}$ un'arbitraria soluzione di (13.27). Allora la matrice dei termini diagonali di $\tilde{M}$ è data da*

$$\mathrm{diag}(\tilde{M}) = \mathrm{diag}\{\frac{1}{\gamma_1}, \frac{1}{\gamma_2}, \dots, \frac{1}{\gamma_n}\} \tag{13.30}$$

Prova Premoltiplicando per $\mathbf{p}_0^T$ primo e secondo membro di $\tilde{M} = \mathbf{1}_n \mathbf{1}_n^T + P\left(\tilde{M} - \mathrm{diag}(\tilde{M})\right)$ si ottiene

$$\mathbf{p}_0^T \tilde{M} \;=\; \mathbf{p}_0^T \mathbf{1}\mathbf{1}^T + \mathbf{p}_0^T P\left(\tilde{M} - \mathrm{diag}(\tilde{M})\right) \;=\; \mathbf{1}^T + \mathbf{p}_0^T\left(\tilde{M} - \mathrm{diag}(\tilde{M})\right)$$

e quindi

$$\mathbf{p}_0^T \mathrm{diag}(\tilde{M}) = [\, \gamma_1 \quad \gamma_2 \quad \dots \quad \gamma_n \,] \begin{bmatrix} \tilde{m}_{11} & & & \\ & \tilde{m}_{22} & & \\ & & \ddots & \\ & & & \tilde{m}_{nn} \end{bmatrix} = [\, 1 \quad 1 \quad \dots \quad 1 \,],$$

da cui segue $\tilde{m}_{ii} = 1/\gamma_i$ per $1 = 1, 2, \dots, n$. ■

Il lemma 13.4.6 fornisce, in particolare, i valori diagonali della matrice delle medie M, ovvero i valori medi del "tempo di ricorrenza" richiesto affinchè la catena irriducibile, inizializzata in un certo stato S_i, lo rivisiti per la prima volta. Tale media è il reciproco della componente i-esima dell'autovettore stocastico di Perron, quindi, se la catena è regolare, il reciproco della probabilità asintotica che la catena si trovi nello stato S_i.
La proposizione 13.4.7, dimostrando che la soluzione dell'equazione lineare (13.27) è unica, garantisce la sua risolubilità con la regola di Cramer (o con metodi equivalenti).

Proposizione 13.4.7 [Unicità della soluzione per l'equazione delle medie]
Ogni soluzione $\tilde{M}$ dell'equazione (13.27) coincide con la matrice delle medie: $\tilde{M} = M$.

Prova Per la Proposizione precedente $\tilde{M}$ e M soddisfano la condizione

$$\mathrm{diag}(\tilde{M}) = \mathrm{diag}(M) = \mathrm{diag}\{ \frac{1}{\gamma_1}, \frac{1}{\gamma_2}, \cdots, \frac{1}{\gamma_n} \}$$

Sottraendo membro a membro le identità

$$\begin{aligned} M &= \mathbf{1}\mathbf{1}^T + P\,(M - \mathrm{diag}(M)) \\ \tilde{M} &= \mathbf{1}\mathbf{1}^T + P\left(\tilde{M} - \mathrm{diag}(\tilde{M})\right) \end{aligned}$$

si ottiene $M - \tilde{M} = P(M - \tilde{M})$.
Per ogni i, la colonna i-esima $\mathbf{v}$ di $M - \tilde{M}$ soddisfa la relazione $\mathbf{v} = P\mathbf{v}$, quindi è nulla o è un autovettore destro di P relativo all'autovalore massimale $\lambda_0 = 1$. Nel secondo caso $\mathbf{v}$ dovrebbe essere proporzionale all'autovettore dominante destro $\mathbf{1}$, ma la sua i-esima componente è nulla. Quindi la colonna $\mathbf{v}$ è identicamente nulla. ■

- Esercizio 13.4.4 [Matrice fondamentale di una catena irriducibile] Si consideri una catena irriducibile $\mathcal{C}$, con matrice di transizione P e autovettore stocastico sinistro $\mathbf{p}_0^T$. Si ponga, come nel caso di una catena regolare $\bar{P} := \mathbf{1}_n \mathbf{p}_0^T$ e si dimostri che
  (i) la matrice $I_n - P + \bar{P}$ è invertibile. L'inversa $Z := (I_n - P + \bar{P})^{-1}$ viene chiamata "matrice fondamentale della catena irriducibile".
  (ii) $Z(I_n - P) = (I_n - P)Z = I_n - \bar{P}$;

  ♯ *Soluzione. Supponiamo che il vettore* $\mathbf{v}$ *appartenga al nucleo di* $\left(I_n - P + \bar{P}\right)$ *e quindi soddisfi*

$$\left(I_n - P + \bar{P}\right)\mathbf{v} = \mathbf{0}. \tag{13.31}$$

  Premoltiplicando (13.31) per $\mathbf{p}_0^T$ *e tenuto conto che* $\mathbf{p}_0^T\left(I_n - P\right) = \mathbf{0}^T$, *si ricava*

$$0 \;=\; \mathbf{p}_0^T \bar{P} \mathbf{v} \;=\; \mathbf{p}_0^T \mathbf{1}_n \mathbf{p}_0^T \mathbf{v} \;=\; \mathbf{p}_0^T \mathbf{v}. \tag{13.32}$$

  Ma allora abbiamo $\bar{P}\mathbf{v} = \mathbf{0}$ *e, di conseguenza,* $(I_n - P)\mathbf{v} = \mathbf{0}$. *Se* $\mathbf{v}$ *non fosse nullo, sarebbe autovettore di* P *relativo all'autovalore* $\lambda_0 = 1$, *quindi* $\mathbf{v}$ *sarebbe proporzionale al vettore* $\mathbf{1}_n$, *e ciò è*

incompatibile con (13.32). Poiché il nucleo di $\left(I_n - P + \bar{P}\right)$ *contiene solo il vettore nullo, la matrice è invertibile.*
(ii) Tenuto conto della idempotenza di $\mathbf{1}_n\mathbf{p}_0^T$ *e della proprietà di assorbimento di* $\mathbf{1}_n\mathbf{p}_0^T$ *rispetto a* P *(esercizio 13.3.2), si ottiene*

$$\left(I_n - P + \mathbf{1}_n\mathbf{p}_0^T\right)(I_n - \mathbf{1}_n\mathbf{p}_0^T) = I_n - \mathbf{1}_n\mathbf{p}_0^T - P + P\mathbf{1}_n\mathbf{p}_0^T + \mathbf{1}_n\mathbf{p}_0^T - \mathbf{1}_n\mathbf{p}_0^T\mathbf{1}_n\mathbf{p}_0^T = I_n - P$$

e premoltiplicando per Z *membro di destra e di sinistra si ricava appunto* $Z(I_n - P) = I_n - \mathbf{1}_n\mathbf{p}_0^T$. *Partendo dal prodotto* $(I_n - \mathbf{1}_n\mathbf{p}_0^T)\left(I_n - P + \mathbf{1}_n\mathbf{p}_0^T\right)$ *si ottiene* $(I_n - P)Z = I_n - \mathbf{1}_n\mathbf{p}_0^T$.

- Esercizio 13.4.5 [Struttura della matrice delle medie] Si consideri una catena irriducibile $\mathcal{C}$, con matrice di transizione P. Siano

  $\mathbf{p}_0^T = [\,\gamma_1 \quad \gamma_2 \quad \dots \quad \gamma_n\,]$ l'autovettore stocastico sinistro e $\bar{P} = \mathbf{1}\mathbf{p}_0^T$,
  $\operatorname{diag}(M) = \operatorname{diag}\{\frac{1}{\gamma_1}, \frac{1}{\gamma_2}, \dots, \frac{1}{\gamma_n}\}$ la matrice diagonale estratta dalla matrice delle medie dei tempi di attesa,
  $Z = (I_n - P + \bar{P})^{-1}$ la matrice fondamentale della catena.

  Posto
  $$\bar{M} := \left(I_n - Z + \mathbf{1}_n\mathbf{1}_n^T\operatorname{diag}(Z)\right)\operatorname{diag}(M) \tag{13.33}$$
  allora
  (i) $\operatorname{diag}(\bar{M}) = \operatorname{diag}(M)$;
  (ii) $\bar{M}$ soddisfa l'equazione (13.27);
  (iii) $\bar{M}$ coincide con la matrice M delle medie dei tempi di attesa f_{ik}.

  ♯ *Soluzione: (i) applicando l'identità* $\operatorname{diag}(A\operatorname{diag}(B)) = \operatorname{diag}(A)\operatorname{diag}(B)$ *si ricava*

  $$\begin{aligned}\operatorname{diag}(\bar{M}) &= \operatorname{diag}\left((I_n - Z + \mathbf{1}_n\mathbf{1}_n^T\operatorname{diag}(Z)\right)\operatorname{diag}(M) = \left(I_n - \operatorname{diag}(Z) + \operatorname{diag}(\mathbf{1}_n\mathbf{1}_n^T)\operatorname{diag}(Z)\right)\operatorname{diag}(M)\\ &= (I_n - \operatorname{diag}(Z) + I_n\operatorname{diag}(Z))\operatorname{diag}(M) = \operatorname{diag}(M)\end{aligned}$$

  (ii) da (13.33) e (i) si ricava

  $$\bar{M} - \operatorname{diag}(\bar{M}) = \bar{M} - \operatorname{diag}(M) = \left(-Z + \mathbf{1}_n\mathbf{1}_n^T\operatorname{diag}(Z)\right)\operatorname{diag}(\bar{M}),$$

  da cui, tenuto conto del punto (ii) dell'esercizio 13.4.4,

  $$\begin{aligned}P(\bar{M} - \operatorname{diag}(\bar{M})) &= \left(-PZ + \mathbf{1}_n\mathbf{1}_n^T\operatorname{diag}(Z)\right)\operatorname{diag}(\bar{M})\\ &= \left(-PZ + \mathbf{1}_n\mathbf{1}_n^T\operatorname{diag}(Z)\right)\operatorname{diag}(\bar{M}) + \bar{M} - \left(I_n - Z + \mathbf{1}_n\mathbf{1}_n^T\operatorname{diag}(Z)\right)\operatorname{diag}(\bar{M})\\ &= \bar{M} - (I_n - Z + PZ)\operatorname{diag}(\bar{M})\\ &= \bar{M} - \operatorname{diag}(\bar{M}) + (I_n - P)Z\operatorname{diag}(\bar{M})\\ &= \bar{M} - \operatorname{diag}(\bar{M}) + (I_n - \mathbf{1}_n\mathbf{p}_0^T)\operatorname{diag}(\bar{M})\\ &= \bar{M} - \mathbf{1}_n\mathbf{p}_0^T\operatorname{diag}(\bar{M})\\ &= \bar{M} - \mathbf{1}_n\mathbf{1}_n^T\end{aligned}$$

  (iii) Per la proposizione 13.4.5 $\bar{M}$ *coincide con la matrice delle medie* M.

- Esercizio 13.4.6 [Matrice fondamentale di una catena regolare] Se $\mathcal{C}$ è una catena regolare con matrice di transizione P e autovettore stocastico sinistro $\mathbf{p}_0^T$, posto $\bar{P} := \lim_{t\to+\infty} P^t = \mathbf{1}_n\mathbf{p}_0^T$
  (i) la matrice $P - \bar{P}$ ha tutti gli autovalori con modulo minore di 1;
  (ii) la matrice fondamentale della catena è esprimibile come

  $$Z = I_n + (P - \bar{P}) + (P - \bar{P})^2 + (P - \bar{P})^3 + \dots$$

  ♯ *Suggerimento: (i) Utilizzando le proprietà di* P *e di* $\bar{P}$ *dimostrate nella proposizione 13.2.2, si ha*

  $$\begin{aligned}(P - \bar{P})^t &= \sum_{\nu=0}^{t-1}\binom{t}{\nu}(-1)^{t-\nu}P^\nu\bar{P}^{t-\nu} + P^t = \sum_{\nu=0}^{t-1}\binom{t}{\nu}(-1)^{t-\nu}P^\nu\bar{P} + P^t\\ &= [\sum_{\nu=0}^{t-1}\binom{t}{\nu}(-1)^{t-\nu}]\bar{P} + P^t = [\sum_{\nu=0}^{t}\binom{t}{\nu}(-1)^{t-\nu} - \binom{t}{t}(-1)^0]\bar{P} + P^t = -\bar{P} + P^t.\end{aligned}$$

  Allora $\lim_{t\to+\infty}(P - \bar{P})^t = -\bar{P} + \lim_{t\to+\infty} P^t = 0$ *e ciò implica* $|\Lambda(P - \bar{P})| < 1$.
  (ii) Si utilizzi il lemma 11.5.6, valutando il risolvente $(\lambda I_n - (P - \bar{P}))^{-1}$ *di* $P - \bar{P}$ *per* $\lambda = 1$.

13.5 Classificazione degli stati e classi di comunicazione

Quando si considera una catena di Markov non irriducibile, le modalità di "visita" degli stati che la compongono possono avere carattere sostanzialmente diverso in corrispondenza alle varie configurazioni possibili per gli zeri nella matrice di transizione, e cioè alla presenza o meno di taluni archi nel grafo orientato associato alla catena. In particolare, come vedremo, esistono sottoinsiemi di stati che, durante l'evoluzione della catena, vengono abbandonati definitivamente, mentre altri sottoinsiemi, una volta raggiunti, non possono più essere abbandonati e al loro interno si dispiega una dinamica tipica delle catene irriducibili.

13.5.1 Accessibilità e comunicazione

Per comprendere l'evoluzione di una catena di Markov governata da una generica matrice di transizione conviene procedere ad una classificazione degli stati che la compongono, dal punto di vista della possibilità o dell'impossibilità di transitare verso altri stati.

Definizione 13.5.1 [ACCESSIBILITÀ] *Diciamo che uno stato S_k è accessibile dallo stato S_i se è positiva la probabilità che la catena, partendo al tempo $t = 0$ dallo stato S_i, visiti in qualche istante $t \geq 0$ lo stato S_k.*

L'accessibilità di S_k da S_i richiede che l'elemento in posizione (i, k) sia positivo in qualche potenza della matrice P, ovvero che sia $[P^t]_{ik} > 0$ per qualche $t \geq 0$. Sul grafo orientato la condizione equivale all'esistenza di almeno un cammino orientato che porti dal vertice S_i al vertice S_k.
La relazione di accessibilità non è simmetrica (l'esistenza di un cammino orientato da S_i a S_k non implica quella di un cammino da S_k a S_i), mentre è riflessiva, quando si convenga di includere fra i cammini quello di lunghezza 0, e transitiva (se da S_i c'è un cammino orientato verso S_k e da S_k uno verso S_j, allora c'è un cammino orientato da S_i a S_j).

Definizione 13.5.2 [COMUNICAZIONE] *Diciamo che due stati S_i e S_k comunicano se S_k è accessibile da S_i e S_i è accessibile da S_k.*

Se S_i comunica con S_k, devono essere positivi l'elemento in posizione (i, k) in qualche potenza t_1-esima e l'elemento in posizione (k, i) in qualche potenza t_2-esima[13] della matrice P. Sul grafo orientato la condizione di comunicazione equivale all'esistenza di almeno un cammino orientato che porti da S_i a S_k e di un altro che porti da S_k a S_i o, equivalentemente, di un ciclo orientato passante per entrambi i vertici.
La relazione di comunicazione è riflessiva, simmetrica e transitiva, quindi è una equivalenza. In quanto tale, induce una partizione degli stati in classi di equivalenza ("classi di comunicazione") disgiunte, la cui unione coincide con l'intero insieme degli stati. Una catena è irriducibile se e solo se è costituita da un'unica classe di comunicazione.

Due stati distinti S_i e S_k possono, in alternativa,

- appartenere alla medesima classe di comunicazione (S_k è accessibile da S_i e S_i è accessibile da S_k)
- appartenere a classi di comunicazione diverse (almeno uno dei due non è accessible dall'altro)

[13] t_2 è in generale diverso da t_1.

A seconda della struttura della catena,

- dagli stati di una classe di comunicazione può essere o non essere possibile visitare stati di altre classi di comunicazione;
- una classe di comunicazione può essere visitata, oppure non può esserlo, da stati appartenenti ad altre classi.

Le situazioni possibili sono schematizzate in figura 13.5.1.

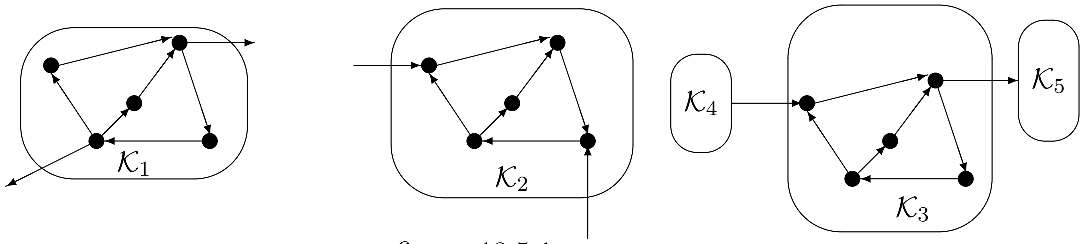


figura 13.5.1

Nel primo caso, dalla classe di comunicazione $\mathcal{K}_1$ si possono visitare stati di altre classi, nel secondo da altre classi di comunicazione è possibile accedere alla classe $\mathcal{K}_2$, nel terzo è possibile sia transitare dalla classe $\mathcal{K}_3$ ad altre classi, sia accedere a $\mathcal{K}_3$ da altre classi. Nel terzo caso, in cui esistono sia stati di $\mathcal{K}_3$ accessibili da un'altra classe $\mathcal{K}_4$, sia stati di un'altra classe $\mathcal{K}_5$ accessibili da $\mathcal{K}_3$, non può esistere un cammino chiuso che esce da $\mathcal{K}_3$ e vi ritorna: se così fosse, esisterebbe una comunicazione fra stati di $\mathcal{K}_3$ e stati di classi diverse, che dovrebbero allora coincidere con $\mathcal{K}_3$.

13.5.2 Classi di comunicazione ergodiche

Definizione 13.5.3 [Classi ergodiche e classi transitorie] *Una classe di comunicazione $\mathcal{K}$ si dice*

- *ergodica*[14] *(o chiusa) se le transizioni da stati di $\mathcal{K}$ verso stati di altre classi hanno probabilità nulla, cioè se nessuno stato fuori di $\mathcal{K}$ è accessibile da (uno stato di) $\mathcal{K}$ (se $\mathcal{K}$ consta di un solo stato, questo viene detto stato "assorbente");*
- *transitoria se sono possibili transizioni da stati di $\mathcal{K}$ verso stati di altre classi, i.e. se esistono stati esterni a $\mathcal{K}$ accessibili da (qualche stato di) $\mathcal{K}$.*

Nella precedente figura 13.5.1, la classe $\mathcal{K}_2$ è ergodica, mentre sono transitorie sia la classe $\mathcal{K}_1$ che la classe $\mathcal{K}_3$.

- Esercizio 13.5.1 Possono esistere classi transitorie con un solo elemento? Quale ne può essere la struttura?

Se ad un certo istante una catena entra in una classe ergodica $\mathcal{K}_e$, essa non potrà più uscirne e da quell'istante la catena si muoverà soltanto entro $\mathcal{K}_e$. In particolare, se la catena raggiunge uno stato assorbente, vi rimarrà indefinitamente. D'altra parte, se una catena entra, o si trova, in una classe transitoria $\mathcal{K}_t$, essa potrà uscirne ma, una volta

[14]la sottomatrice della matrice di transizione P relativa agli stati di una classe ergodica è una matrice irriducible (perché?). Si noti che, mentre una classe ergodica è una particolare classe di comunicazione, non lo sono invece, quando siano in numero maggiore di uno, le classi cicliche che la compongono e che rappresentano una particolare partizione in sottoinsiemi della classe di comunicazione ergodica.

uscita, non potrà più rientrarvi. Ove ciò si verificasse, esisterebbe un cammino γ che parte da uno stato $S_i \in \mathcal{K}_t$, passa per almeno uno stato $S_k \notin \mathcal{K}_t$ e termina in uno stato $S_f \in \mathcal{K}_t$.
Se $S_i = S_f$, è immediato che S_k e S_i comunicano e quindi dovrebbero appartenere alla stessa classe di comunicazione; se $S_i \neq S_f$, esisterebbe un cammino entro la classe $\mathcal{K}_t$ che connette S_f a S_i e, combinandolo con γ, esisterebbe ancora una comunicazione fra S_k e S_i.

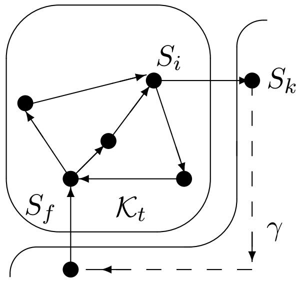


figura 13.5.2

Proposizione 13.5.4 [Esistenza e accessibilità delle classi ergodiche] *Ogni catena $\mathcal{C}$ possiede almeno una classe ergodica. Se uno stato S_i di $\mathcal{C}$ non appartiene ad una classe ergodica, esiste almeno un cammino orientato con inizio in S_i che termina in una classe ergodica.*

Prova Se $\mathcal{C}$ ha una sola classe di comunicazione, questa è necessariamente ergodica. Se ne ha più d'una e una di esse, $\mathcal{K}_{i_1}$, è transitoria, esiste un cammino orientato con origine in $\mathcal{K}_{i_1}$ che entra in una classe diversa $\mathcal{K}_{i_2}$. Se anche questa non è ergodica, da $\mathcal{K}_{i_2}$ il cammino può proseguire in una classe $\mathcal{K}_{i_3}$ diversa $\mathcal{K}_{i_2}$, e così via. Poichè le classi di comunicazione sono in numero finito, il cammino dovrà entrare finalmente in una classe ergodica oppure in qualche momento visiterà nuovamente una delle classi transitorie già incontrate. Questa seconda alternativa implica che esista una comunicazione fra stati appartenenti a classi di comunicazione distinte (fra i quali per definizione non può esistere comunicazione): una contraddizione. Quindi $\mathcal{C}$ ha almeno una classe ergodica.
La seconda affermazione si dimostra in modo analogo. Dallo stato S_i in una classe transitoria esiste un cammino che porta in una classe diversa da quella di S_i, e se questa non è ergodica, si può proseguirlo in un'altra classe diversa dalla seconda etc. La mancata visitazione di almeno uno stato in una classe ergodica consentirebbe di tornare in una delle classi transitorie già incontrate e comporterebbe la comunicazione fra due stati appartenenti a classi distinte. ■

Proposizione 13.5.5 [Appartenenza asintotica alle classi ergodiche] *Per $t \to +\infty$ è unitaria la probabilità che lo stato di una catena di Markov $\mathcal{C}$ appartenga a una classe ergodica.*

Prova Se lo stato iniziale appartiene a una classe ergodica, non c'è nulla da provare, dato che con probabilità 1 la catena visiterà solo gli stati della classe.
Si è dimostrato nella proposizione precedente che se lo stato iniziale S_i appartiene a una classe transitoria, esiste almeno un cammino che porta da S_i in una classe ergodica. Indichiamo con

- μ_i il numero minimo di passi richiesto per entrare da S_i in una classe ergodica,
- $\bar{p}_i < 1$ la probabilità che la catena, inizializzata in S_i, rimanga nelle classi transitorie nei primi μ_i passi

e poniamo

$$\bar{\mu} := \max_{\substack{i: S_i \text{ in classe} \\ \text{transitoria}}} \{\mu_i\}, \qquad \bar{p} := \max_{\substack{i: S_i \text{ in classe} \\ \text{transitoria}}} \{\bar{p}_i\} < 1$$

Partendo da qualsiasi stato in una classe transitoria, dopo $\bar{\mu}$ passi la probabilità che la catena sia in una classe transitoria è minore o eguale a $\bar{p}$. Se la catena al $\bar{\mu}$-esimo passo è in una classe transitoria, la probabilità di rimanere nelle classi transitorie nei successivi $\bar{\mu}$ passi è ancora non superiore a $\bar{p}$: quindi la probabilità di rimanervi nei primi $2\bar{\mu}$ passi è non superiore a $\bar{p}^2$. Induttivamente, la probabilità di rimanere nelle classi transitorie nei primi $k\bar{\mu}$ passi non supera $\bar{p}^k$, per ogni $k \geq 1$, ovvero la probabilità di entrare in una classe ergodica nei primi $k\bar{\mu}$ passi è almeno $1 - \bar{p}^k$. Essa tende a 1 quando $k \to +\infty$. ■

Ordiniamo gli stati di una catena di Markov $\mathcal{C}$ secondo il seguente criterio:

a) per ciascuna classe di comunicazione, gli indici dei suoi stati formano un intervallo di $\mathbb{Z}$;

b) [LIVELLO ZERO] le classi ergodiche precedono tutte le classi transitorie;

c) fra le classi transitorie, vengono elencate

   [PRIMO LIVELLO] per prime quelle in cui ogni transizione verso le classi ergodiche si realizza direttamente (i.e. senza passare attraverso altre classi transitorie)

   [SECONDO LIVELLO] poi le classi transitorie per cui il numero massimo di classi transitorie (inclusa la classe di partenza) incontrate per raggiungere le classi ergodiche è 2;

   [TERZO LIVELLO] poi le classi transitorie per cui il numero massimo di classi transitorie (inclusa la classe di partenza) incontrate per raggiungere le classi ergodiche è 3;

   etc.

   Si noti che da nessuna classe del livello i-esimo si può accedere ad altre classi del medesimo livello o dei livelli superiori: altrimenti esisterebbe un percorso dal livello i-esimo che raggiunge le classi ergodiche incontrando un numero di classi transitorie superiore a i.

La figura 15.5.3 illustra il criterio di precedenza fra le classi.

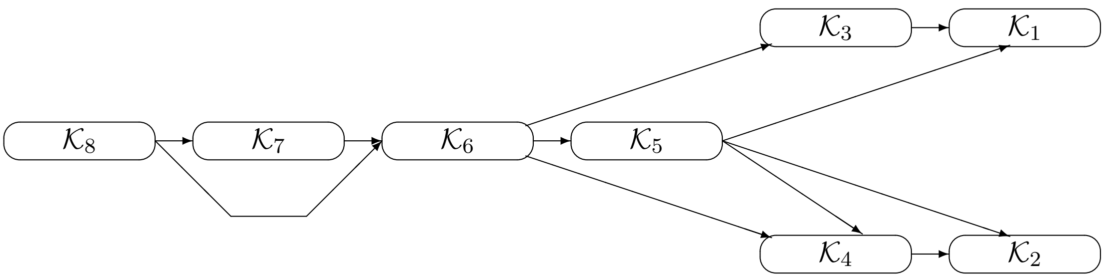


Figura 13.5.3

La matrice di transizione P, una volta ordinati gli stati e le classi di comunicazione secondo la procedura indicata, assume forma triangolare a blocchi. Nel caso specifico dell'esempio di figura 13.5.3, indicando con il nome delle classi i blocchi diagonali (i cui elementi sono le probabilità di transizione entro le classi) e con il simbolo "+" i blocchi fuori diagonale in cui figurano alcuni elementi positivi, si perviene ad una matrice di transizione con la seguente struttura a blocchi (i primi due blocchi diagonali si riferiscono alle transizioni entro le due classi ergodiche):

$$P = \begin{bmatrix} \mathcal{K}_1 & & | & & & O & & & \\ 0 & \mathcal{K}_2 & | & & & & & & \\ - & - & - & - & - & - & - & - & - \\ + & 0 & | & \mathcal{K}_3 & & & & & \\ 0 & + & | & 0 & \mathcal{K}_4 & & & O & \\ + & + & | & 0 & + & \mathcal{K}_5 & & & \\ 0 & 0 & | & + & + & + & \mathcal{K}_6 & & \\ 0 & 0 & | & 0 & 0 & 0 & + & \mathcal{K}_7 & \\ 0 & 0 & | & 0 & 0 & 0 & + & + & \mathcal{K}_8 \end{bmatrix}$$

Si noti che i blocchi diagonali

- se di dimensione 1×1 si riferiscono ad una classe di comunicazione ergodica con un solo stato (assorbente), e in tal caso sono unitari, oppure a una classe di comunicazione transitoria con un solo stato, e in tal caso possono assumere qualsiasi valore non negativo e minore di 1;
- se si riferiscono a classi di comunicazione con almeno due stati, sono irriducibili e stocastici o substocastici[15] a seconda che la classe sia ergodica o transitoria. L'irriducibilità segue dal fatto che entro una classe di comunicazione è positiva la probabilità di transitare da qualsiasi suo stato S_i a qualsiasi suo stato S_j in un congruo numero di passi; quindi, scelta una posizione (i, j) nel blocco diagonale, esiste una opportuna potenza del blocco in cui l'elemento in posizione (i, j) è positivo.

Lo studio di una catena $\mathcal{C}$ costituita soltanto da classi ergodiche consegue immediatamente dalla discussione delle catene irriducibili e, in particolare, di quelle regolari. Se tutte le classi sono ergodiche, la matrice di transizione, a meno di una rinumerazione degli stati che soddisfi il criterio del punto (a), ha struttura diagonale a blocchi

$$P = \begin{bmatrix} P_1 & & & \\ & P_2 & & \\ & & \ddots & \\ & & & P_h \end{bmatrix}. \tag{13.34}$$

Siamo in presenza di h catene irriducibili giustapposte e non interagenti, con matrici di transizione $P_1, P_2, \ldots, P_h$ irriducibili. Se, ad esempio, lo stato iniziale appartiene alla prima classe, o se la distribuzione di probabilità iniziale è concentrata solo sulla prima classe, le altre classi non saranno mai visitate e lo studio di $\mathcal{C}$ si riduce a quello della catena caratterizzata da P_1.

- Esercizio 13.5.2 Si discuta il comportamento di una catena $\mathcal{C}$, priva di classi transitorie e costituita da $h > 1$ classi ergodiche comprendenti rispettivamente $\nu_1, \nu_2, \ldots, \nu_h$ stati, quando la probabilità iniziale è positiva su stati di classi diverse, e quindi assume la forma $\mathbf{x}^T(0) = [\,\mathbf{x}_1^T \;\; \mathbf{x}_2^T \;\; \cdots \;\; \mathbf{x}_h^T\,]$, con $\mathbf{x}_i^T = \left[\,\pi_1^{(i)} \;\; \ldots \;\; \pi_{\nu_i}^{(i)}\,\right]$, $\sum_{i,j} \pi_j^{(i)} = 1$ e almeno due blocchi $\mathbf{x}_i^T$ e $\mathbf{x}_j^T$, $i \neq j$, positivi.

[15] Una matrice quadrata nonnegativa è (propriamente) substocastica se la somma degli elementi di ciascuna delle sue righe non supera 1 e per una riga almeno la somma è minore di 1.

13.6 Catene con classi transitorie

Quando sono presenti in una catena $\mathcal{C}$ una o più classi transitorie e si adotta l'ordinamento degli stati discusso nel precedente paragrafo, la matrice di transizione ha struttura a blocchi

$$P = \begin{bmatrix} E & 0 \\ R & Q \end{bmatrix} \tag{13.35}$$

Il blocco E, di dimensione $r \times r$, riguarda la dinamica nelle classi ergodiche, che comprendono complessivamente r stati, ed è la matrice stocastica (13.34).
Il blocco Q, di dimensione $(n-r) \times (n-r)$ è una matrice substocastica[16], poichè alcune sue righe (quelle in corrispondenza alle quali non sono nulle le righe di R) hanno somma minore di 1. Più precisamente, ogni blocco diagonale irriducibile della forma normale di Q ha somme di riga non superiori ad 1 e almeno una riga a somma minore di 1 (altrimenti il blocco irriducibile farebbe riferimento ad una classe ergodica). Perciò gli autovalori di ciascun blocco diagonale nella forma normale di Q hanno modulo minore di 1 e Q^t converge a zero al crescere di t. Nella potenza t-esima di P

$$P^t = \begin{bmatrix} E^t & 0 \\ \star & Q^t \end{bmatrix}$$

l'elemento $[Q^t]_{i,j}$ di Q^t dà la probabilità che la catena, partendo dallo stato iniziale S_i in una classe transitoria, si trovi dopo t passi nello stato S_j delle classi transitorie. Il fatto che $[Q^t]_{i,j}$ converga a zero al crescere di t corrisponde alla proprietà, già evidenziata, che qualunque sia lo stato iniziale S_i di una classe transitoria, la probabilità che la catena si trovi all'istante t in qualche classe ergodica tende a 1 al divergere di t e quindi tende a zero la probabilità che la catena si trovi in una classe transitoria. Applicando il lemma 10.5.6 sulla serie di C.Neumann, si conclude che la serie $I + Q + Q^2 + \dots$ converge all'inversa di $I - Q$, e tale inversa è una matrice positiva.

Definizione 13.6.1 [MATRICE FONDAMENTALE DELLE CLASSI TRANSITORIE] *Se la catena $\mathcal{C}$ ha classi transitorie e (13.35) ne è la matrice di transizione, la matrice fondamentale delle classi transitorie è la matrice positiva*

$$L := (I_{n-r} - Q)^{-1} = I_{n-r} + Q + Q^2 + \dots \tag{13.36}$$

Introduciamo, per ogni coppia di stati S_i e S_j nelle classi transitorie, la variabile casuale v_{ij} che fornisce il numero di volte che la catena visita S_j partendo da S_i. Se $s_{ij}^{(t)}$, $t = 0, 1, 2, \dots$ denota la variabile casuale

$$s_{ij}^{(t)} = \begin{cases} 1 & \text{se, assumendo } S(0) = S_i, \text{ la catena all'istante } t \text{ visita } S_j \\ 0 & \text{se, assumendo } S(0) = S_i, \text{ la catena all'istante } t \text{ non visita } S_j, \end{cases}$$

si ha $v_{ij} = \sum_{t=0}^{\infty} s_{ij}^{(t)}$, e la media della variabile v_{ij} è data da

$$\begin{aligned} \mathbf{E}(v_{ij}) &:= \mathbf{E}\Big(\sum_{t=0}^{\infty} s_{ij}^{(t)}\Big) = \sum_{t=0}^{\infty} \mathbf{E}(s_{ij}^{(t)}) \\ &= \sum_{t=0}^{\infty} \{1 \cdot [P^t]_{ij} + 0 \cdot (1 - [P^t]_{ij})\} = \sum_{t=0}^{\infty} [P^t]_{ij} = \sum_{t=0}^{\infty} [Q^t]_{ij}, \end{aligned} \tag{13.37}$$

[16] Gli elementi di Q e delle sue potenze prendono gli indici nell'insieme $\{r+1, r+2, \dots, n\}$, i.e. fra gli indici degli stati delle classi transitorie. Quindi, se $i > r$ e $j > r$, per ogni $k > 0$ si ha $[P^k]_{ij} = [Q^k]_{ij}$.

dove si sono mantenti per gli elementi del blocco Q gli indici di riga e di colonna ereditati da P. Abbiamo così provato la prima parte della seguente

Proposizione 13.6.2 [Numero di visite agli stati transitori] *L'elemento* (i, j) *della matrice fondamentale* $L = (I_{n-r} - Q)^{-1}$ *fornisce il numero medio delle volte che la catena* $\mathcal{C}$ *visita lo stato transitorio* S_j*, a partire dallo stato transitorio* $S(0) = S_i$*.*
La componente i*-esima del vettore colonna* $L\mathbf{1}_{n-r}$ *rappresenta il numero medio complessivo di passi che la catena, partendo dallo stato transitorio* i*-esimo, compie nelle classi transitorie, prima di entrare in una delle classi ergodiche.*

Prova Per la seconda parte basta osservare che la prima componente del vettore $L\mathbf{1}_{n-r}$ è somma delle componenti della prima riga di L, quindi somma
- del numero medio di volte che la catena visita il primo stato transitorio
- più il numero medio di volte che visita il secondo
- ....
- più il numero medio di volte che visita l'$(n-r)$-esimo,

quando lo stato iniziale è il primo degli stati transitori. Analogo discorso vale per le altre componenti di $L\mathbf{1}_{n-r}$. ■

Mentre la matrice fondamentale L dà informazioni sul comportamento della catena finchè lo stato rimane nelle classi transitorie, la matrice LR fornisce informazione circa le modalità di passaggio dalle classi transitorie alle classi ergodiche. Si ha infatti

Proposizione 13.6.3 [Probabilità di prima visita ad una classe ergodica] *Indichiamo con* b_{ik} *la probabilità che la catena, partendo dallo stato transitorio* S_i*, entri in una delle classi ergodiche visitando per primo lo stato* S_k *di tale classe. Si ha allora*

$$b_{ik} = [LR]_{ik} \tag{13.38}$$

Prova Se l'entrata della catena in una classe ergodica avviene visitando come primo stato ergodico lo stato S_k, il passaggio da S_i a S_k può richiedere un passo, con probabilità p_{ik}, oppure più passi, con la preliminare visitazione di alcuni stati transitori. Questa seconda eventualità comporta, al primo passo, il passaggio da S_i verso uno qualsiasi degli stati transitori S_j, con probabilità p_{ij}, e il successivo cammino da S_j verso S_k, in un congruo numero di passi e con probabilità b_{jk}.
Quindi le probabilità b_{ik} sono espresse in forma implicita dalle relazioni

$$b_{ik} = p_{ik} + \sum_{j:S_j \text{ nelle classi transitorie}} p_{ij} b_{jk} \tag{13.39}$$

con p_{ik} elemento di R e p_{ij} elementi di Q.
In termini matriciali, posto $B = [b_{ik}]$, da (13.39) si ricava $B = R + QB$ e quindi

$$B = (I_{n-r} - Q)^{-1} R = LR.$$ ■

- Esercizio 13.6.1 Nella matrice $B \in \mathbb{R}^{(n-r)\times r}$ è unitaria la somma degli elementi di ciascuna riga.
♯ *Suggerimento: Partendo da* S_i*, la prima visita delle classi ergodiche si verifica secondo* r *modalità, esuastive e mutuamente esclusive, a seconda che il primo stato ergodico visitato sia* $S_1, S_2 \ldots$ *o* S_r*. In alternativa, si utilizzi l'identità* $R\mathbf{1}_r = \mathbf{1}_{n-r} - Q\mathbf{1}_{n-r}$ *ottenendo (si giustifichino i passaggi)*

$$\begin{aligned} B\mathbf{1}_r &= (R + QR + Q^2R + \ldots)\mathbf{1}_r = R\mathbf{1}_r + Q(\mathbf{1}_{n-r} - Q\mathbf{1}_r) + Q^2(\mathbf{1}_{n-r} - Q\mathbf{1}_r) + \ldots \\ &= R\mathbf{1}_r + Q\mathbf{1}_{n-r} = \mathbf{1}_{n-r} \end{aligned}$$

Esempio 13.6.1 [LINEA DI PRODUZIONE] Il processo di manifattura di un prodotto consiste di tre stadi successivi, che chiameremo S_5, S_4, S_3.
Una unità di prodotto entra inizialmente nello stadio di lavorazione S_5 , alla fine della lavorazione viene ispezionata e

- con probabilità p **p**assa allo stadio di lavorazione successivo S_4,
- con probabilità s viene **s**cartata e non più recuperata,
- con probabilità r viene **r**iciclata nello stadio S_5.

Analoga procedura, e con le stesse probabilità, vale per le unità di prodotto che sono entrate nello stadio S_4. Per quelle entrate nello stadio S_3, conclusa la lavorazione

- con probabilità p escono dalla linea di produzione e **p**assano alla commercializzazione S_2,
- con probabilità s vengono **s**cartate e non più recuperate,
- con probabilità r vengono **r**iciclate nello stadio S_3.

Se denotiamo con S_1 lo stadio (unico) in cui entra l'unità di prodotto scartata, il processo può essere schematizzato dalla seguente catena di Markov,

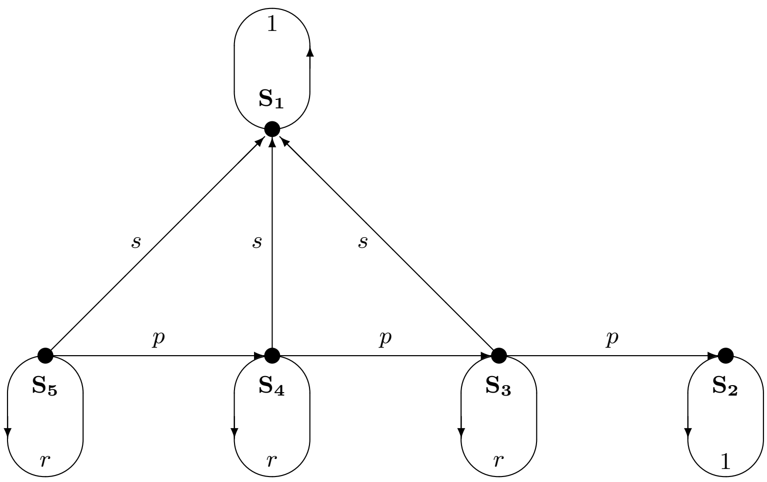


Figura 13.6.1

che ha matrice di transizione

$$P = \begin{bmatrix} 1 & 0 & | & 0 & 0 & 0 \\ 0 & 1 & | & 0 & 0 & 0 \\ - & - & | & - & - & - \\ s & p & | & r & 0 & 0 \\ s & 0 & | & p & r & 0 \\ s & 0 & | & 0 & p & r \end{bmatrix} = \begin{bmatrix} E & 0 \\ R & Q \end{bmatrix}.$$

La catena ha due stati assorbenti, S_1 e S_2, e tre classi transitorie, ciascuna comprendente un solo stato: $\{S_3\}, \{S_4\}, \{S_5\}$.
La matrice fondamentale delle classi transitorie è data da $L = (I_3 - Q)^{-1} = (I_3 - rI_3 - pN)^{-1}$, dove con N si è indicata la matrice nilpotente

$$N = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Si ha così

$$L \quad = \quad ((1-r)I_3 - pN)^{-1} = \frac{1}{1-r}\left(I_3 - \frac{p}{1-r}N\right)^{-1}$$

$$= \frac{1}{1-r}\Big(I_3 + \frac{p}{1-r}N + (\frac{p}{1-r})^2 N^2\Big) = \begin{bmatrix} \frac{1}{1-r} & 0 & 0 \\ \frac{p}{(1-r)^2} & \frac{1}{1-r} & 0 \\ \frac{p^2}{(1-r)^3} & \frac{p}{(1-r)^2} & \frac{1}{1-r} \end{bmatrix}$$

$$L\mathbf{1}_3 = \begin{bmatrix} \frac{1}{1-r} & 0 & 0 \\ \frac{p}{(1-r)^2} & \frac{1}{1-r} & 0 \\ \frac{p^2}{(1-r)^3} & \frac{p}{(1-r)^2} & \frac{1}{1-r} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{1-r} \\ \frac{p+(1-r)}{(1-r)^2} \\ \frac{p^2+p(1-r)+(1-r)^2}{(1-r)^3} \end{bmatrix}$$

$$B = LR = \begin{bmatrix} \frac{s}{1-r} & \frac{p}{1-r} \\ \frac{sp+s(1-r)}{(1-r)^2} & \frac{p^2}{(1-r)^2} \\ \frac{sp^2+sp(1-r)+s(1-r)^2}{(1-r)^3} & \frac{p^3}{(1-r)^3} \end{bmatrix} = \begin{bmatrix} 1-\frac{p}{1-r} & \frac{p}{1-r} \\ 1-\frac{p^2}{(1-r)^2} & \frac{p^2}{(1-r)^2} \\ 1-\frac{p^3}{(1-r)^3} & \frac{p^3}{(1-r)^3} \end{bmatrix}$$

L'ultima riga della matrice fondamentale L fornisce i valori medi del numero di volte che un'unità di prodotto rimane nei tre stadi di lavorazione: $1/1-r$ nello stadio iniziale S_5, $p/(1-r)^2$ nello stadio intermedio S_4, $p^2/(1-r)^3$ nell'ultimo stadio di lavorazione S_3.

È possibile, in questo caso piuttosto semplice, reinterpretare il risultato: r^k è la probabilità che una unità di prodotto che entra in lavorazione nello stadio S_5 vi venga riciclata per (almeno!) k volte; $r^k - r^{k+1}$ è la probabilità che vi venga riciclata per esattamente k volte, ossia che S_5 sia visitato esattamente $k+1$ volte. Il numero medio di visite nello stadio S_5 è dato allora da

$$1(1-r) + 2(r-r^2) + 3(r^2-r^3) + \ldots = 1 + (2r-r) + (3r^2-2r^2) + \ldots = 1 + r + r^2 + \ldots = \frac{1}{1-r}$$

Per ottenere il numero medio di visite compiute nello stadio S_4 da un'unità che entra nello stadio iniziale S_5, dobbiamo anzitutto considerare la probabilità che in qualche momento l'unità passi da S_5 a S_4; questa è data dalla somma delle probabilità che il passaggio avvenga in un passo, in due passi (al primo passo un riciclo e al secondo il passaggio), in tre passi (nei primi due passi un riciclo, al terzo il passaggio) etc., ovvero da

$$p + rp + r^2p + r^3p + \ldots = p/(1-r). \tag{13.40}$$

(Si noti che (13.40) rappresenta anche, nella matrice B, la probabilità che, quando il processo si trova nello stadio finale di lavorazione S_3, lo stato assorbente in cui esso finalmente entra sia lo stato S_2, di commercializzazione.)

A questo punto, la probabilità che un'unità di prodotto che entra in lavorazione nello stadio S_5 venga riciclata nello stadio S_4 per esattamente k volte è data dal prodotto della probabilità di passaggio da S_5 in S_4 per la probabilità che, una volta giunta in S_4, l'unità vi venga riciclata esattamente k volte, ovvero da $\frac{p}{1-r}(r^k - r^{k+1})$. Ragionando come in precedenza, si ottiene il penultimo elemento dell'ultima riga di L.

Infine, è facile convincersi che la probabilità che in qualche momento la catena passi in S_3, partendo inizialmente da S_5 è il quadrato di $p/(1-r)$, da cui si ottiene facilmente il primo elemento dell'ultima riga di L.

- **Esempio 13.6.2** [Torneo con due giocatori] Vogliamo modellare una situazione in cui due giocatori $\mathcal{A}$ e $\mathcal{B}$ hanno inizialmente a disposizione a e b unità di conto (monete, fiches, etc.), rispettivamente. Durante ciascuna mano di gioco, il giocatore $\mathcal{A}$ ha probabilità p di vincere, $p \in (0,1)$, il giocatore $\mathcal{B}$ ha probabilità $q = 1-p$. Il giocatore che vince la mano guadagna da quello che perde una unità di conto. Il gioco ha termine quando un giocatore ha vinto tutte le $a+b=c$ unità di conto disponibili (e quindi l'altro rimane senza più nulla da giocare).

Il gioco può essere visto come una catena di Markov, in cui gli stati corrispondono al numero di unità di conto disponibili al giocatore $\mathcal{A}$. Il grafo che rappresenta la dinamica del processo in un passo è il seguente (il pedice dello stato rappresenta qui la somma disponibile al giocatore $\mathcal{A}$):

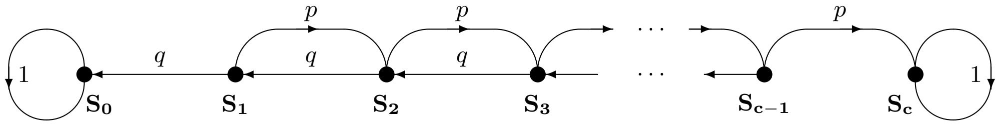


Figura 13.6.2

La catena, con due stati assorbenti S_0 e S_c e un'unica classe transitoria $\{S_1, S_2, \ldots, S_{c-1}\}$, è la prima delle "passeggiate casuali" considerate nell'esempio 13.2.1. Conviene, prima di scrivere la matrice di transizione, rinumerare gli stati, riservando gli indici più bassi agli stati assorbenti. Si ha allora

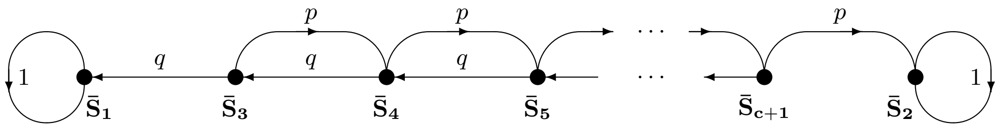


Figura 13.6.3

e la corrispondente matrice di transizione è

$$P = \left[\begin{array}{cc|cccccc} 1 & 0 & & & O & & & \\ 0 & 1 & & & & & & \\ \hline q & 0 & 0 & p & 0 & & & \\ 0 & 0 & q & 0 & p & 0 & & \\ 0 & 0 & 0 & q & 0 & p & & \\ & & \vdots & & \ddots & \ddots & \ddots & \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & p \\ 0 & p & 0 & 0 & 0 & \ldots & q & 0 \end{array}\right] = \begin{bmatrix} E & 0 \\ R & Q \end{bmatrix}$$

Come sappiamo, la matrice $B = (I - Q)^{-1}R$ fornisce le probabilità di entrare nelle due classi ergodiche partendo da uno qualsiasi dei $c - 1$ stati della classe transitoria: la prima colonna di B contiene le probabilità che il giocatore $\mathcal{A}$ perda (i.e. entri nella prima classe ergodica), la seconda colonna di B le probabilità che il giocatore $\mathcal{A}$ vinca, quando parte da una disponibilità iniziale di $1, 2, \ldots, c - 1$ unità di conto. Possiamo valutare direttamente le probabilità di vittoria di $\mathcal{A}$ (i.e. la seconda colonna di B) con una procedura ricorsiva che sfrutta la struttura particolare di Q. Se indichiamo con $\xi_1, \xi_2, \ldots, \xi_{c-1}$ le componenti incognite della seconda colonna di B, per determinarle dobbiamo risolvere l'equazione

$$(I - Q)\begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{c-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ p \end{bmatrix} \tag{13.41}$$

ovvero il sistema lineare non omogeneo di $c - 1$ equazioni in $c - 1$ incognite

$$\left\{\begin{array}{ccccccc} & & \xi_1 & - & p\xi_2 & = & 0 \\ -q\xi_1 & + & \xi_2 & - & p\xi_3 & = & 0 \\ -q\xi_2 & + & \xi_3 & - & p\xi_4 & = & 0 \\ \ldots & & & & & & \\ -q\xi_{c-3} & + & \xi_{c-2} & - & p\xi_{c-1} & = & 0 \\ -q\xi_{c-2} & + & \xi_{c-1} & & & = & p \end{array}\right. \tag{13.42}$$

Se introduciamo le grandezze ausilarie ξ_0 e ξ_c e le vincoliamo a valere rispettivamente 0 e 1, (13.42) assume struttura omogenea e uniforme in tutte le equazioni

$$\begin{cases} -q\xi_0 & + & \xi_1 & - & p\xi_2 & = & 0 \\ -q\xi_1 & + & \xi_2 & - & p\xi_3 & = & 0 \\ -q\xi_2 & + & \xi_3 & - & p\xi_4 & = & 0 \\ \dots \\ -q\xi_{c-3} & + & \xi_{c-2} & - & p\xi_{c-1} & = & 0 \\ -q\xi_{c-2} & + & \xi_{c-1} & - & p\xi_c & = & 0 \end{cases}$$

e le componenti ξ_i sono legate dalla formula ricorsiva

$$p\xi_{i+1} = \xi_i - q\xi_{i-1} \ . \tag{13.43}$$

Posto $\mathbf{x}(i) = \begin{bmatrix} \xi_i \\ \xi_{i-1} \end{bmatrix}$, da (13.43) si ricavano le equazioni di un sistema lineare osservabile e in evoluzione libera

$$\begin{aligned} \mathbf{x}(i+1) &= \begin{bmatrix} 1/p & -q/p \\ 1 & 0 \end{bmatrix} \mathbf{x}(i) = F\mathbf{x}(i) \\ \xi_i &= \begin{bmatrix} 1 & 0 \end{bmatrix} \mathbf{x}(i) = H\mathbf{x}(i) \end{aligned}$$

con autovalori 1 e q/p, del quale si hanno informazioni parziali sullo stato iniziale e sullo stato finale

$$\mathbf{x}(1) = \begin{bmatrix} \xi_1 \\ 0 \end{bmatrix}, \quad \mathbf{x}(c) = \begin{bmatrix} 1 \\ \xi_{c-1} \end{bmatrix}.$$

L'uscita è combinazione lineare dei modi:

– se $q \neq p$, i modi sono $\{1^i : i \in \mathbb{N}\}$ e $\{(q/p)^i : i \in \mathbb{N}\}$ e abbiamo

$$\xi_i = k_1 (q/p)^i + k_2.$$

Imponendo le condizioni $0 = \xi_0 = k_1 + k_2$ e $1 = \xi_c = k_1 (q/p)^c + k_2$, si determinano le costanti

$$k_1 = \frac{1}{(q/p)^c - 1}, \quad k_2 = -k_1$$

Quindi la probabilità che il giocatore $\mathcal{A}$ vinca, partendo da una situazione iniziale in cui $\mathcal{A}$ e $\mathcal{B}$ possiedono rispettivamente a e $b = c - a$ fiches, è data da

$$\xi_a = \frac{1}{(q/p)^c - 1}(q/p)^a - \frac{1}{(q/p)^c - 1} = \frac{(q/p)^a - 1}{(q/p)^c - 1} \tag{13.44}$$

– se $q = p = 1/2$ (quindi se il gioco è equo), l'autovalore 1 ha molteplicità 2 e i modi del sistema sono le successioni $\{1^i : i \in \mathbb{N}\}$ e $\{\binom{i}{1} : i \in \mathbb{N}\}$. In questo caso abbiamo

$$\xi_i = k_1 \binom{i}{1} + k_2$$

e imponendo le condizioni $0 = \xi_0 = k_1 0 + k_2$ e $1 = \xi_c = k_1 \binom{c}{1} + k_2$, si determinano le costanti

$$k_1 = \frac{1}{c}, \quad k_2 = 0.$$

La probabilità che il giocatore $\mathcal{A}$ vinca è data da $\xi_a = \frac{1}{c}\binom{a}{1} = \frac{a}{c}$, come si ottiene anche da (13.44) passando al limite quando q/p tende a 1.

13.7 Cenno ai modelli genetici

L'informazione genetica nella materia vivente è contenuta nei *cromosomi*, corpuscoli a struttura sequenziale contenuti nel nucleo della cellula. Lungo ciascun cromosoma c'è una progressione di *locazioni (= "loci")*, ciascuna delle quali è occupata da un *gene*, specifico per quel locus e responsabile, da solo o con altri geni presenti in loci diversi, per un carattere elementare dell'organismo (il colore degli occhi, la rugosità dei semi, etc). La variabilità dei caratteri si riconnette al fatto che ogni gene può presentare diverse varianti: ciascuna di esse è detta un *allele* del gene, e il gene che occupa effettivamente un locus lungo il cromosoma rappresenta quindi una scelta all'interno di un insieme di possibili alleli.
Per molti organismi viventi, le cellule (eccettuate quelle deputate alla riproduzione, i *gameti*, di cui diremo poi) contengono un corredo cromosomico *diploide*: i cromosomi della cellula sono presenti in *coppie omologhe*: ognuno dei due cromosomi della coppia contiene una progressione di loci che codifica la stessa sequenza di caratteri, arrangiata nel medesimo ordine. Due loci corrispondenti lungo i cromosomi omologhi sono occupati dal medesimo gene, ma è possibile che nei due loci il gene sia presente con alleli diversi, appartenenti alla varietà specifica per quel gene.

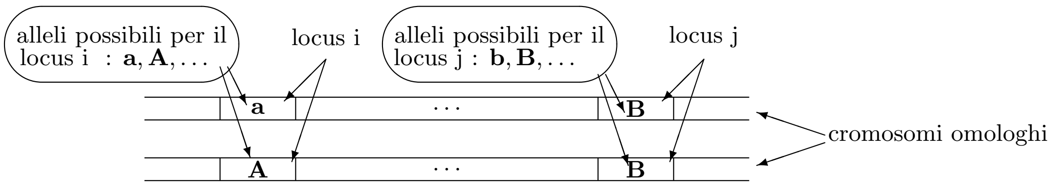


figura 13.7.1

La diploidia non è una caratteristica universale del corredo cromosomico delle specie viventi. Una cellula che contiene una sola copia di ciascun cromosoma viene detta *aploide*. Vi sono specie viventi, in particolare quelle costituite da organismi dalla struttura meno evoluta (muschi, funghi, etc.), i cui cicli di vita includono stadi di sviluppo in cui tutte le cellule sono aploidi. Tuttavia, anche gli organismi diploidi producono alcune cellule aploidi, i gameti per la riproduzione sessuata.
In termini estremamente semplificati, la riproduzione sessuata comporta la fusione del materiale genetico proveniente da due organismi genitori . Ciascuno di essi produce cellule aploidi (i gameti: uova o spermi) mediante un processo di divisione cellulare, detto meiosi.

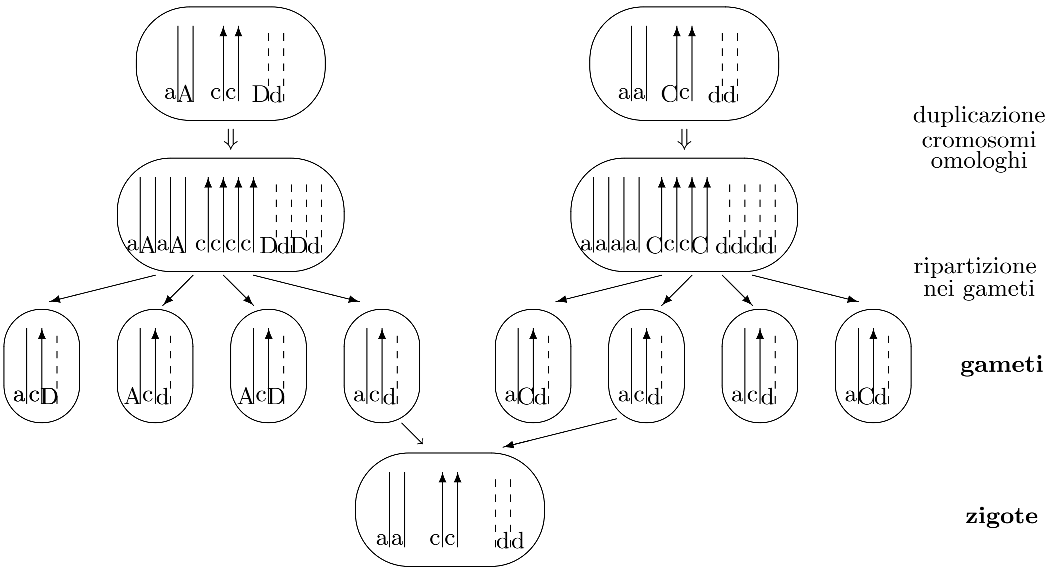


Figura 13.7.2 Il corredo genetico di ciascun genitore comprende tre coppie di cromosomi omologhi e su ciascuna coppia è indicato un locus particolare, occupabile da due alleli diversi di un gene.

Nella schematizzazione più elementare, e con riferimento a una specifica coppia di cromosomi omologhi,
- ciascun genitore raddoppia entrambi i cromosomi omologhi, producendo una cellula che contiene una quaterna di cromosomi omologhi;
- le 4 copie omologhe si ripartiscono in 4 cellule progenitrici (gameti) , ciascuna delle quali contiene quindi una sola copia di ciascun cromosoma , e quindi un solo allele per ciascun gene;
- la riproduzione sessuale si verifica quando due gameti, provenienti ciascuno da uno dei genitori, fondono in un'unica cellula diploide, lo *zigote*, che darà poi luogo all'individuo maturo.

- **Esempio 13.7.1** [Modello di Wright-Fisher] Il modello genetico di Wright-Fischer fa riferimento a una popolazione diploide di individui ermafroditi[17], in cui ciascun individuo può accoppiarsi con qualsiasi individuo della popolazione. Ci porremo nelle condizioni più semplici, in assenza di mutazioni cromosomiche e di meccanismi selettivi che attribuiscano vantaggi riproduttivi a individui con un corredo cromosomico particolare. Ci limiteremo inoltre al caso di un carattere determinato da un singolo locus ℓ, per il quale siano possibili più alleli: uno specifico, che indicheremo con la lettera A, e uno o più altri, che indicheremo collettivamente con l'indeterminata X. Vogliamo costruire un modello che renda conto dell'evoluzione temporale del numero ν_A degli alleli A complessivamente presenti in una popolazione di N individui, ossia della frequenza relativa $\nu_A/2N$ dell'allele A sull'intero pool degli alleli A e X presenti nel locus ℓ.

  Per ricavare le equazioni del modello, supponiamo inoltre che
  - le generazioni successive non si sovrappongano, nel senso che non è possibile l'accoppiamento fra individui di generazioni diverse;
  - la popolazione mantenga un livello costante N in ogni generazione;
  - ciascun individuo della generazione $(t+1)$-esima sia frutto dell' accoppiamento di due individui della generazione t-esima, ciascuno dei quali dona al figlio un solo gamete, recante con eguale probabilità uno dei due alleli (non necessariamente distinti) presenti nel locus del genitore, oppure sia figlio di un unico individuo, che dona al figlio due gameti, ciascuno dei quali reca, con eguale probabilità, uno dei due alleli presenti nel locus del genitore.

  (i) Con queste premesse, il meccanismo che porta alla creazione del corredo genetico nel locus ℓ degli N individui nella generazione $t+1$-esima è equivalente a quello di ripetere per $2N$ volte l'estrazione di un campione da un'urna che contiene i $2N$ alleli presenti nel locus ℓ della generazione t-esima, reimbussolando dopo ogni estrazione il campione estratto: le prime due estrazioni forniranno il corredo genetico del primo individuo, le seconde due quelle del secondo, etc. Se nella generazione t-esima il pool genetico degli N individui contiene, nel locus ℓ, i alleli A su un totale di $2N$, in ciascuna estrazione, la probabilità di estrarre un gamete che contiene l'allele A è $i/2N$ e la probabilità di estrarne complessivamente j nelle $2N$ estrazioni corrisponde al termine della distribuzione binomiale

$$p_{ij} = \binom{2N}{j} \left[\frac{i}{2N}\right]^j \left[\frac{2N-i}{2N}\right]^{2N-j} \tag{13.45}$$

  Se caratterizziamo lo stato della popolazione alla generazione t-esima con il numero degli alleli A presenti negli N individui che la costituiscono, gli stati sono $2N+1$, ovvero $\{0, 1, 2, \ldots, 2N\}$ e da una generazione alla successiva la probabilità di transizione è espressa, appunto, dalla matrice

$$P = [p_{ij}] = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ [\frac{2N-1}{2N}]^{2N} & \binom{2N}{1}[\frac{1}{2N}][\frac{2N-1}{2N}]^{2N-1} & \binom{2N}{2}[\frac{1}{2N}]^2[\frac{2N-1}{2N}]^{2N-2} & \ldots & [\frac{1}{2N}]^{2N} \\ [\frac{2N-2}{2N}]^{2N} & \binom{2N}{1}[\frac{2}{2N}][\frac{2N-2}{2N}]^{2N-1} & \binom{2N}{2}[\frac{2}{2N}]^2[\frac{2N-2}{2N}]^{2N-2} & \ldots & [\frac{2}{2N}]^{2N} \\ [\frac{2N-3}{2N}]^{2N} & \binom{2N}{1}[\frac{3}{2N}][\frac{2N-3}{2N}]^{2N-1} & \binom{2N}{2}[\frac{3}{2N}]^2[\frac{2N-3}{2N}]^{2N-2} & \ldots & [\frac{3}{2N}]^{2N} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ [\frac{1}{2N}]^{2N} & \binom{2N}{1}[\frac{2N-1}{2N}][\frac{1}{2N}]^{2N-1} & \binom{2N}{2}[\frac{2N-1}{2N}]^2[\frac{1}{2N}]^{2N-2} & \ldots & [\frac{2N-1}{2N}]^{2N} \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}$$

[17] In natura tale situazione si verifica piuttosto spesso, soprattutto per le specie vegetali.

i cui elementi sono dati dalla (13.45). È chiaro che la catena non è irriducibile e che le classi di comunicazione sono $\{0\}, \{1, 2, \ldots, 2N-1\}, \{2N\}$, ergodiche la prima e l'ultima, transitoria l'altra.

(ii) In corrispondenza all'autovalore $\lambda_0 = 1$, la matrice possiede l'autovettore destro strettamente positivo $\mathbf{1}_{2N+1}$ e un autovettore destro positivo, linearmente indipendente dal primo,

$$\mathbf{v}_0 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 2N \end{bmatrix}. \tag{13.46}$$

Per verificarlo, basta ricorrere all'identità $k\binom{2N}{k} = 2N\binom{2N-1}{k-1}$, $\quad k = 1, 2, \ldots, 2N$ e osservare che $P\mathbf{v}_0$ ha come riga di indice i

$$\begin{aligned} \sum_{k=0}^{2N} k\binom{2N}{k}\left[\frac{i}{2N}\right]^k\left[\frac{2N-i}{2N}\right]^{2N-k} &= \sum_{k=1}^{2N} 2N\frac{i}{2N}\binom{2N-1}{k-1}\left[\frac{i}{2N}\right]^{k-1}\left[\frac{2N-i}{2N}\right]^{2N-k} \\ &= i\sum_{k=1}^{2N}\binom{2N-1}{k-1}\left[\frac{i}{2N}\right]^{k-1}\left[\frac{2N-i}{2N}\right]^{2N-1-(k-1)} = i \end{aligned} \tag{13.47}$$

Per gli sviluppi successivi, converrà riscrivere (13.47) nella forma seguente:

$$\frac{i}{2N} = \left\{\sum_{k=1}^{2N-1}\binom{2N}{k}\left[\frac{i}{2N}\right]^k\left[\frac{2N-i}{2N}\right]^{2N-k}\frac{k}{2N}\right\} + \left[\frac{i}{2N}\right]^{2N} \tag{13.48}$$

(iii) Permutando gli stati e riportando dapprima le due classi ergodiche $\{0\}$ e $\{2N\}$ e poi la classe transitoria, la matrice di transizione assume la forma già discussa all'inizio del paragrafo, ovvero

$$\begin{array}{ccccccc} \text{stati} \rightarrow & 0 & 2N & & 1 & \ldots & 2N-1 \end{array}$$

$$\begin{bmatrix} E & 0 \\ R & Q \end{bmatrix} = \left[\begin{array}{cc|ccc} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \hline [\frac{2N-1}{2N}]^{2N} & [\frac{1}{2N}]^{2N} & \binom{2N}{1}[\frac{1}{2N}][\frac{2N-1}{2N}]^{2N-1} & \ldots & \binom{2N}{2N-1}[\frac{1}{2N}]^{2N-1}[\frac{2N-1}{2N}] \\ [\frac{2N-2}{2N}]^{2N} & [\frac{2}{2N}]^{2N} & \binom{2N}{1}[\frac{2}{2N}][\frac{2N-2}{2N}]^{2N-1} & \ldots & \binom{2N}{2N-1}[\frac{2}{2N}]^{2N-1}[\frac{2N-2}{2N}] \\ [\frac{2N-3}{2N}]^{2N} & [\frac{3}{2N}]^{2N} & \binom{2N}{1}[\frac{3}{2N}][\frac{2N-3}{2N}]^{2N-1} & \ldots & \binom{2N}{2N-1}[\frac{3}{2N}]^{2N-1}[\frac{2N-3}{2N}] \\ \ldots & \ldots & \ldots & & \\ [\frac{1}{2N}]^{2N} & [\frac{2N-1}{2N}]^{2N} & \binom{2N}{1}[\frac{2N-1}{2N}][\frac{1}{2N}]^{2N-1} & \ldots & \binom{2N}{2N-1}[\frac{2N-1}{2N}]^{2N-1}[\frac{1}{2N}] \end{array}\right]$$

Nella nuova base gli autovettori stocastici sinistri corrispondenti all'autovalore $\lambda_0 = 1$ sono, com'è ovvio, $[1 \quad 0 \quad 0 \quad \ldots \quad 0]$ e $[0 \quad 1 \quad 0 \quad \ldots \quad 0]$.

(iv) Asintoticamente, la popolazione entrerà in uno dei due stati ergodici, corrispondenti ad avere presenti solo individui privi dell'allele A, oppure ad avere solo individui omozigoti per l'allele A. Le probabilità $b_{i,0}$ o $b_{i,2N}$ che la catena, inizializzata con i alleli A, entri, rispettivamente, nello stato ergodico in cui sono presenti solo alleli X o in quello caratterizzato dalla presenza del solo allele A, sono elementi di

$$B = \begin{bmatrix} b_{1,0} & b_{1,2N} \\ b_{2,0} & b_{2,2N} \\ \vdots & \vdots \\ b_{2N-1,0} & b_{2N-1,2N} \end{bmatrix} = (I_{2N-1} - Q)^{-1}R. \tag{13.49}$$

Se riscriviamo la seconda colonna di (13.49) nella forma

$$\begin{bmatrix} b_{1,2N} \\ b_{2,2N} \\ \vdots \\ b_{2N-1,2N} \end{bmatrix} = Q \begin{bmatrix} b_{1,2N} \\ b_{2,2N} \\ \vdots \\ b_{2N-1,2N} \end{bmatrix} + \begin{bmatrix} [\frac{1}{2N}]^{2N} \\ [\frac{2}{2N}]^{2N} \\ \vdots \\ [\frac{2N-1}{2N}]^{2N} \end{bmatrix}$$

e teniamo conto di (13.48), è immediato ottenere

$$\begin{bmatrix} b_{1,2N} \\ b_{2,2N} \\ \vdots \\ b_{2N-1,2N} \end{bmatrix} = \begin{bmatrix} \frac{1}{2N} \\ \frac{2}{2N} \\ \vdots \\ \frac{2N-1}{2N} \end{bmatrix}$$

Quindi una popolazione con una presenza iniziale di i alleli A ha probabilità $b_{i,2N} = \frac{i}{2N}$ di diventare omozigote per A (e quindi di entrare nel secondo stato ergodico della catena), mentre $b_{i,0} = \frac{2N-i}{2N}$ rappresenta la probabilità dell'evento complementare, che il gene A sparisca dalla popolazione.

(v) Consideriamo ora la variabile casuale a valori interi $\nu_A(t)$ che rappresenta il numero di alleli A presenti nella popolazione al tempo t. Tenendo conto dell'espressione (13.46) per $\mathbf{v}_0$, il valore atteso di $\nu_A(t)$ è dato da

$$E[\nu_A(t)] = \sum_{k=0}^{2N} k \operatorname{pr}\{\nu_A(t) = k\} = \mathbf{x}^T(t)\mathbf{v}_0$$

All'istante successivo, abbiamo

$$E[\nu_A(t+1)] = \sum_{k=0}^{2N} k \operatorname{pr}\{\nu_A(t+1) = k\} = \mathbf{x}^T(t+1)\mathbf{v}_0 = \mathbf{x}^T(t)P\mathbf{v}_0 = \mathbf{x}^T(t)\mathbf{v}_0 = E[\nu_A(t)], \quad (13.50)$$

quindi il valore atteso della variabile ν_A rimane costante nel tempo. In particolare, se all'istante iniziale sono presenti nella popolazione i alleli A,

$$E[\nu_A(0)] = \sum_{k=0}^{2N} k \operatorname{pr}\{\nu_A(0) = k\} = \mathbf{x}^T(0)\mathbf{v}_0 = i$$

e in tutti gli istanti successivi il valore medio del numero di alleli A rimane costantemente i. Il risultato è in accordo anche con il comportamento asintotico: per grandi valori di t la catena è nello stato 0 con probabilità $\frac{2N-i}{2N}$ e nello stato $2N$ con probabilità $\frac{i}{2N}$: con un lieve abuso di notazione, il valor medio della variabile è

$$\begin{aligned} E[\nu_A(\infty)] &= \sum_{k=0}^{2N} k \operatorname{pr}\{\nu_A(\infty) = k\} \\ &= 0 \operatorname{pr}\{\nu_A(\infty) = 0\} + 2N \operatorname{pr}\{\nu_A(\infty) = 2N\} = 2N\frac{i}{2N} = i\,. \end{aligned}$$

- **Esempio 13.7.2** [Modello di Moran] (i) Il modello genetico di Moran fa riferimento a una popolazione $\mathcal{P}$ di N individui, che si riproduce in via asessuata, per semplice duplicazione. Ciascun individuo può avere nel proprio genotipo uno di due alleli[18], A o a. Il meccanismo secondo cui evolve la popolazione è il seguente. Nell'intervallo $(t, t+1)$, in successione
  - dalla popolazione $\mathcal{P}$ viene estratto a caso un "genitore", che si replica, dando luogo a una copia esatta di se stesso;
  - il genitore estratto viene "reimbussolato";

[18]Il modello può applicarsi, più generalmente, all'evoluzione di una popolazione costituita da individui di due tipi, prescindendo dall'interpretazione genetica.

- da $\mathcal{P}$ viene estratto a caso un secondo individuo, che muore;
- il nuovo nato viene aggiunto alla popolazione, ripristinando il livello N.

Gli stati possibili della sistema al tempo t sono $N+1$, tanti quanti sono gli individui con allele A. Se al tempo t lo stato è i (ovvero ci sono i individui con allele A e $N-i$ con allele a), all'istante $t+1$ sono possibili tre stati, se $0 < i < N$:

- $i+1$, corrispondente alla scelta di un genitore con allele A e alla sua duplicazione, nonché alla scelta di un individuo con allele a e alla sua morte;
- i, corrispondente alla scelta di un genitore con allele A e alla sua duplicazione, nonché alla scelta e alla morte di un individuo con allele A, oppure alla scelta di un genitore con allele a e alla sua duplicazione, nonché alla scelta e alla morte di un individuo con allele a;
- $i-1$, corrispondente alla scelta di un genitore con allele a e alla sua duplicazione, nonché alla scelta e alla morte di un individuo con allele A.

Le tre eventualità hanno probabilità

$$p_{i,i+1} = \frac{i}{N}\frac{N-i}{N} = \frac{i(N-i)}{N^2},\ p_{i,i} = (\frac{i}{N})^2+(\frac{N-i}{N})^2 = \frac{i^2+(N-i)^2}{N^2},\ p_{i,i-1} = \frac{N-i}{N}\frac{i}{N} = \frac{i(N-i)}{N^2}$$

È un facile esercizio verificare che le formule precedenti valgono anche se $i=0$ o $i=N$ e che, per ogni i compreso fra 0 e N si ha

$$p_{i,i-1}(i-1) + p_{i,i}i + p_{i,i+1}(i+1) = i \tag{13.51}$$

La matrice di transizione è tridiagonale:

$$P = \frac{1}{N^2}\begin{bmatrix} N^2 & & & & & & \\ (N-1) & 1+(N-1)^2 & (N-1) & & & & \\ 0 & 2(N-2) & 2^2+(N-2)^2 & 2(N-2) & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & (N-2)2 & (N-2)^2+2^2 & (N-2)2 & 0 \\ & & & & (N-1) & (N-1)^2+1 & (N-1) \\ & & & & & & N^2 \end{bmatrix}$$

Anche in questo caso ci sono due stati assorbenti, 0 e N, e una classe transitoria $\{1,2,\dots,N-1\}$.

(ii) In corrispondenza all'autovalore $\lambda_0 = 1$, la matrice P possiede l'autovettore destro strettamente positivo $\mathbf{1}_{2N+1}$ e, ricorrendo a (13.51), si verifica che

$$\mathbf{v}_0 = \begin{bmatrix} 0 \\ 1 \\ 2 \\ \vdots \\ N \end{bmatrix}.$$

è un altro autovettore destro positivo, indipendente dal primo.

Permutando gli stati, in modo che compaiano al primo posto quelli ergodici, si perviene alla matrice di transizione

$$\begin{bmatrix} E & 0 \\ R & Q \end{bmatrix} = \left[\begin{array}{cc|cccccc} 1 & 0 & & & & & & \\ 0 & 1 & & & & & & \\ \hline p_{1,0} & 0 & p_{1,1} & p_{1,2} & & & & \\ 0 & 0 & p_{2,1} & p_{2,2} & p_{2,3} & & & \\ & & & & & & & \\ \vdots & \vdots & & \ddots & \ddots & \ddots & & \\ & & & & & & & \\ 0 & 0 & & & p_{N-2,N-3} & p_{N-2,N-2} & p_{N-2,N-1} & \\ 0 & p_{N-1,N} & & & & p_{N-1,N-2} & p_{N-1,N-1} & \end{array}\right]$$

(iii) Se indichiamo con $b_{i,0}$ e $b_{i,N}$ le probabilità che la catena, inizializzata con i alleli A, entri nello stato ergodico sorrispondente all'assenza dell'allele A o in quello corrispondente alla presenza esclusiva dell'allele A, la seconda colonna della matrice B soddisfa l'equazione

$$Q \begin{bmatrix} b_{1,N} \\ b_{2,N} \\ \vdots \\ b_{N-1,N} \end{bmatrix} = \begin{bmatrix} b_{1,N} \\ b_{2,N} \\ \vdots \\ b_{N-1,N} \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \vdots \\ p_{N-1,N} \end{bmatrix}. \tag{13.52}$$

Poichè da (13.51) abbiamo

$$\begin{aligned} Q \begin{bmatrix} 1 \\ 2 \\ \vdots \\ N-2 \\ N-1 \end{bmatrix} &= \begin{bmatrix} p_{1,1} & p_{1,2} & & & & \\ p_{2,1} & p_{2,2} & p_{2,3} & & & \\ & \ddots & \ddots & \ddots & & \\ & & & p_{N-2,N-3} & p_{N-2,N-2} & p_{N-2,N-1} \\ & & & & p_{N-1,N-2} & p_{N-1,N-1} \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ \vdots \\ N-2 \\ N-1 \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ 2 \\ \vdots \\ N-2 \\ N-1 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ N\,p_{N-1,N} \end{bmatrix}, \end{aligned} \tag{13.53}$$

dividendo entrambi i membri di (13.53) per N, otteniamo

$$Q \begin{bmatrix} 1/N \\ 2/N \\ \vdots \\ (N-2)/N \\ (N-1)N \end{bmatrix} = \begin{bmatrix} 1/N \\ 2/N \\ \vdots \\ (N-2)/N \\ (N-1)/N \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ p_{N-1,N} \end{bmatrix} \tag{13.54}$$

La soluzione di (13.52) è unica (perché?), quindi, per confronto con (13.54) possiamo concludere che $b_{i,N} = \frac{i}{N},\ i = 0, 1, \ldots, N-1$. Procedendo in modo analogo, si ricava $b_{i,0} = \frac{N-i}{N}$.

- Esercizio 13.7.1 Si verifichi che nel modello di Moran la variabile casuale $\nu_A(t)$, che fornisce il numero di alleli presenti nella popolazione al tempo t, ha un valore atteso $E[\nu_A(t)]$ che non dipende da t, ma soltanto dalla distribuzione iniziale di probabilità $\mathbf{x}^T(0)$.

- Esercizio 13.7.2 [Modello di Moran con mutazioni] Si supponga che, nella fase riproduttiva del modello di Moran, un indivuo con allele A possa duplicarsi
  - con probabilità $0 < \alpha < 1$ in un individuo con il medesimo allele,
  - con probabilità $1 - \alpha$, per effetto di una mutazione, in un individuo con allele a.

  Similmente, un individuo con l'allele a potrà duplicarsi con probabilità $0 < \beta < 1$ in un individuo con allele a e con probabilità $1 - \beta$ in un individuo con allele A.
  (i) Si determinino $p_{i,i+1}, p_{i,i}$ e $p_{i,i-1}$ e la matrice di transizione quando sono presenti le mutazioni.
  (ii) Si verifichi che la presenza di mutazioni dà luogo a una catena di Markov regolare. Qual è l'interpretazione biologica di questo fatto?