Catene di Markov

A.A. 2008–2008

Prof. Carlo Sempi
Dipartimento di Matematica
"Ennio De Giorgi"
Università del Salento
carlo.sempi@unisalento.it

13 marzo 2009

Indice

Capitolo 1

Introduzione alle catene di Markov discrete

1.1 Definizione di catena di Markov

Si dice *processo aleatorio* sopra uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ un'applicazione $X : \Omega \times \mathbf{T} \to \mathbf{R}$, ove $\mathbf{T} \subset \mathbf{R}$, tale che, per ogni $t \in \mathbf{T}$, la funzione $\omega \to X(\omega, t)$ risulti una v.a. . Di solito è $\mathbf{T} = \mathbf{R}$ oppure $\mathbf{T} = \mathbf{R}_+$, o, ancora, $\mathbf{T} = \mathbf{N}$ (o $\mathbf{T} = \mathbf{Z}_+$); $\mathbf{T}$ si chiama *insieme dei tempi.* Se $\mathbf{T} = \mathbf{R}$ (oppure $\mathbf{T} = \mathbf{R}_+$) si dice che il processo è *a tempo continuo*; se invece $\mathbf{T} = \mathbf{N}$ o $\mathbf{T} = \mathbf{Z}_+$, esso si dice *a tempo discreto.* Nel seguito considereremo solo quest'ultimo caso. Un processo aleatorio a tempo discreto è, essenzialmente, una successione di v.a. definite sullo spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$ mediante $X_n(\omega) := X(\omega, n)$. Un tale processo si dice *indipendente* se, e solo se, tali sono le v.a. della successione.

Considereremo ora una classe particolare di processi aleatorî a tempo discreto: supporremo che ogni v.a.del processo assuma valori nell'insieme finito

$$S := \{s_1, s_2, \dots, s_r\}.$$

Un tale processo si dice *finito*; S si dirà *l'insieme degli stati* e i suoi elementi $s_1, s_2, \dots, s_r$ si diranno *stati.* Se $X_n = s_j$, si dirà, usualmente, che, al tempo $t = n$, il processo si trova nello stato s_j; non è difficile abituarsi a questo linguaggio.

Se il processo è indipendente, risulta

$$\mathbb{P}\left[X_{n+1} = s^{(n+1)} \mid X_0 = s^{(0)}, X_1 = s^{(1)}, \dots, X_n = s^{(n)}\right] = \mathbb{P}\left(X_{n+1} = s^{(n+1)}\right)$$

oppure, in generale, $\mathbb{P}\left[X_{n+1} = s^{(n+1)} \mid A_n\right] = \mathbb{P}\left(X_{n+1} = s^{(n+1)}\right)$, ove $s^{(j)}$ denota lo stato nel quale il processo si trova al tempo $t = j$ e A_n è un qualsiasi evento che dipende solo dai valori assunti dal processo sino al tempo $t = n$, cioè dipende solo dai valori delle v.a. X_1, X_2,..., X_n. Cosí, in un processo indipendente, la legge di X_{n+1} non dipende dalla storia precedente del processo.

Nel caso generale, lo stato del processo al tempo $t = n + 1$ dipenderà da tutta la storia precedente; perciò, la classe di processi piú semplice da studiare dopo i processi indipendenti è costituita da quei processi per i quali lo stato, al tempo $t = n + 1$, dipenda dallo stato al tempo $t = n$, ma non dalla storia precedente l'istante $t = n$.

Definizione 1.1.1. Si dice *processo di* MARKOV *(finito)* un processo aleatorio tale che risulti, per ogni $n \in \mathbf{Z}_+$

$$\mathbb{P}\left[X_{n+1} = s^{(n+1)} \mid X_0 = s^{(0)}, X_1 = s^{(1)}, \dots, X_n = s^{(n)}\right] \\ = \mathbb{P}\left[X_{n+1} = s^{(n+1)} | X_n = s^{(n)}\right]$$

oppure

$$\mathbb{P}\left[X_{n+1} = s^{(n+1)} \mid X_n = s^{(n)}, A_{n-1}\right] = \mathbb{P}\left[X_{n+1} = s^{(n+1)} \mid X_n = s^{(n)}\right].$$

Questa proprietà si dice *di* MARKOV e si chiamano *probabilità di transizione al tempo* $t = n$ le probabilità condizionate

$$p_{ij}(n) := \mathbb{P}\left(X_n = s_j \mid X_{n-1} = s_i\right) \qquad (i, j = 1, 2, \dots, r).$$

Si dice, invece, *catena di* MARKOV *(finita)* un processo di MARKOV (finito) per il quale risulti $p_{ij}(n) = p_{ij}$ per ogni $n \in \mathbf{N}$, per il quale, cioè, le probabilità di transizione non dipendano dal tempo. Si dice *matrice di transizione* della catena di MARKOV la matrice $r \times r$ data da $\Pi = (p_{ij})$ e si dice *vettore delle probabilità iniziali* il vettore riga $p(0) = (p_1^0, p_2^0, \dots, p_r^0)$ ove $p_i^0 := \mathbb{P}(X_0 = s_i)$. Il vettore delle probabilità iniziali dà, evidentemente, la distribuzione di probabilità dei varî stati all'istante iniziale. In generale, si dice *vettore di probabilità* un vettore riga con componenti positive e di somma eguale a 1.

Conoscendo il vettore $p(0)$ e la matrice di transizione Π si può calcolare la probabilità $\mathbb{P}(X_n = s_i)$ che al tempo $t = n$ il processo si trovi nello stato s_i:

$$\begin{aligned}\mathbb{P}(X_n = s_j) &= \sum_{i=1}^{r} \mathbb{P}(X_n = s_j | X_{n-1} = s_i)\, \mathbb{P}(X_{n-1} = s_i) \\ &= \sum_{i=1}^{r} p_{ij}\, \mathbb{P}(X_{n-1} = s_i)\end{aligned}$$

che, in notazione matriciale, indicato con $p(n)$ il vettore (riga)

$$p(n) := \left(\mathbb{P}(X_n = s_1), \mathbb{P}(X_n = s_2), \dots, \mathbb{P}(X_n = s_r)\right),$$

si scrive $p(n) = p(n-1)\,\Pi$. Ripetute applicazioni di questa relazione danno

$$p(n) = p(0)\,\Pi^n. \tag{1.1.1}$$

La matrice di transizione Π è una matrice *stocastica*, tale, cioè, che tutti i suoi elementi siano positivi e che sia eguale a 1 la somma di ogni riga:

$$\sum_{j=1}^{r} p_{ij} = \sum_{j=1}^{r} \mathbb{P}\left(X_1 = s_j \mid X_0 = s_i\right) = 1 \quad (i = 1, 2, \dots, r).$$

Noti il vettore delle probabilità $p(0) = (p_1, p_2, \dots, p_r)$ iniziali e la matrice di transizione $\Pi = (p_{ij})$, si può calcolare la legge congiunta delle v.a. X_1, X_2,..., X_n.

Infatti, usando ripetutamente la proprietà di MARKOV, si ha

$$\begin{aligned}
\mathbb{P}(X_0 &= s(0), X_1 = s(1), \dots, X_n = s(n)) \\
&= \mathbb{P}(X_1 = s(1), \dots, X_n = s(n) \mid X_0 = s(0))\, \mathbb{P}(X_0 = s(0)) \\
&= p_{s(0)}\, \mathbb{P}(X_1 = s(1) \mid X_0 = s(0)) \\
&\qquad \mathbb{P}(X_2 = s(2), \dots, X_n = s(n) \mid X_1 = s(1), X_0 = s(0)) \\
&= p_{s(0)}\, p_{s(0),s(1)}\, \mathbb{P}(X_2 = s(2), \dots, X_n = s(n) \mid X_1 = s(1)) \\
&= \dots = p_{s(0)}\, p_{s(0),s(1)}\, p_{s(1),s(2)} \cdots p_{s(n-1),s(n)}.
\end{aligned} \tag{1.1.2}$$

È immediato convincersi che la (1.1.2) equivale alla proprietà di MARKOV.

Vale il seguente importante teorema, la cui dimostrazione si può far discendere da un fondamentale teorema di KOLMOGOROV che sta alla base della teoria dei processi aleatorî, ma che qui non si dimostrerà.

Teorema 1.1.1. *Siano* Π *una matrice stocastica* $r \times r$ *e* $p = (p_1, p_2, \dots, p_r)$ *un vettore di probabilità. Esistono allora uno spazio di probabilità* $(\Omega, \mathcal{F}, \mathbb{P})$ *ed una catena di* MARKOV *che ha* Π *come matrice di transizione e* p *come vettore delle probabilità iniziali.*

Una catena di MARKOV è una successione di v.a. (X_n) con le proprietà che abbiamo appena introdotto. Tuttavia, solitamente, si parla di una catena di MARKOV come se questa fosse la matrice di transizione Π e il vettore delle probabilità iniziali. L'ultimo teorema giustifica questo atteggiamento.

Se si sa che $X_0 = s_i$, cioè che $p_i^0 = 1, p_j^0 = 0 \quad (j \neq i)$, per la (1.1.1) il vettore $p(n)$ è semplicemente l'i–esima riga della matrice Π^n. Perciò, in generale, è

$$p_{ij}^{(n)} = \mathbb{P}(X_n = s_j | X_0 = s_i).$$

Esempio 1.1.1. (Passeggiata aleatoria con pareti assorbenti). In questo e nei successivi esempî si considera una particella che può occupare un numero finito di posizioni (stati) lungo un segmento e che ad ogni istante si muove di un passo a destra con probabilità p oppure di un passo a sinistra con probabilità $q = 1 - p$. Il diverso comportamento agli estremi (detti *pareti* o *barriere*) determinerà diversi tipi di catena di MARKOV. Si considererà il caso di 5 stati. Se il processo rimane in s_1 o in s_5 una volta che abbia raggiunto uno di questi due stati, la matrice di transizione è

$$\begin{vmatrix} 1 & 0 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 0 & 0 & 1 \end{vmatrix}.$$

Esempio 1.1.2. (Passeggiata aleatoria con pareti riflettenti). La particella ritorna al punto di partenza quando raggiunge uno degli estremi. (La particella "rimbalza"):

$$\begin{vmatrix} 0 & 1 & 0 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & 0 & 1 & 0 \end{vmatrix}.$$

Esempio 1.1.3. Quando la particella giunge in s_1 o in s_5, va in s_3:

$$\begin{vmatrix} 0 & 0 & 1 & 0 & 0 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 0 & 0 & 1 & 0 & 0 \end{vmatrix}.$$

Esempio 1.1.4. Quando la particella giunge in un estremo, ha egual probabilità di rimanere sulla parete che ha raggiunto o di ricomparire sull'altra parete:

$$\begin{vmatrix} 1/2 & 0 & 0 & 0 & 1/2 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 1/2 & 0 & 0 & 0 & 1/2 \end{vmatrix}.$$

Esempio 1.1.5. (Passeggiata aleatoria con pareti periodiche). Quando la particella raggiunge un estremo, ricompare all'altro estremo:

$$\begin{vmatrix} 0 & 0 & 0 & 0 & 1 \\ q & 0 & p & 0 & 0 \\ 0 & q & 0 & p & 0 \\ 0 & 0 & q & 0 & p \\ 1 & 0 & 0 & 0 & 0 \end{vmatrix}.$$

Esempio 1.1.6. (1.10) Esempio Un giocatore può scegliere tra due diversi giochi d'azzardo, che saranno gli stati s_1 e s_2: nel primo vince con probabilità p_1, nel secondo con probabilità p_2. Egli continua con lo stesso gioco se perde, mentre passa all'altro quando vince. La matrice di transizione è

$$\begin{vmatrix} 1-p_1 & p_1 \\ p_2 & 1-p_2 \end{vmatrix}.$$

Al variare di p_1 e di p_2 in $[0,1]$, quest'esempio fornisce tutte le catene di Markov con due stati.

1.2 Classificazione di stati e catene

Per una data catena di Markov, e, quindi, per un'assegnata matrice di transizione Π, si definisca come segue una relazione T in $S \times S$: $s_i \, T \, s_j$ significa

(a) che $s_i = s_j$ oppure

(b) che esiste un numero naturale n tale che $p_{ij}^{(n)} > 0$.

In altre parole, uno stato s_i è nella relazione T con lo stato s_j, se i due stati coincidono oppure se è possibile, in un numero finito di passi, che il processo vada dallo stato s_i allo stato s_j. Se $s_i \, T \, s_j$, si dice che lo stato s_j è *accessibile* da s_i.

Si controlla immediatamente che T è transitiva e riflessiva. Basta controllare che T sia transitiva: sia $s_i T s_j$ e $s_j T s_k$; esistono dunque due numeri naturali m e n

tali che $p_{ij}^{(m)} > 0$ e $p_{jk}^{(n)} > 0$. Nell'identità $\Pi^{m+n} = \Pi^m \Pi^n$ si scriva ora l'elemento di indici i e k:

$$p_{ik}^{(m+n)} = \sum_{h=1}^{r} p_{ih}^{(m)}\, p_{hk}^{(n)} \geq p_{ij}^{(m)}\, p_{jk}^{(n)} > 0,$$

ove la prima diseguaglianza discende dall'essere positivi tutti gli elementi delle matrici in questione; dunque $s_i T s_k$. È pure immediato controllare che se $s_i T s_j \wedge s_j T s_i$ allora $s_i = s_j$ oppure esistono due naturali m e n tali che $p_{ij}^{(n)} > 0$ e $p_{ji}^{(m)} > 0$. Perciò è possibile che in un numero finito di passi il processo vada dallo stato s_i allo stato s_j e, viceversa, dallo stato s_j allo stato s_i. Si osservi che non si richiede che il numero di passi sia lo stesso in entrambi i versi. Si crea cosí una partizione dell'insieme S degli stati in sottoinsiemi, detti *classi*, ogni classe essendo composta dagli stati che comunicano tra loro, vale a dire dagli stati che possono essere raggiunti da ogni stato della classe e che dai quali si possono raggiungere tutti glistati della classe. Poiché gli stati sono in numero finito esisteranno classi (una o piú) nelle quali si può entrare ma non tornare indietro. Tali classi sono dette *ergodiche*; si dicono *ergodici* gli stati di una classe ergodica. Le classi che non sono ergodiche, come pure gli stati che le compongono, si dicono *transienti*.

Ogni catena di Markov (finita) ha almeno una classe ergodica. Vi possono, però, non essere classi transienti: ciò accade se la catena è costituita da un'unica classe ergodica, oppure da piú classi ergodiche, da ognuna delle quali è allora impossibile raggiurne un'altra. In quest'ultimo caso si hanno, di fatto, piú catene di Markov non interagenti. Una volta che il processo abbia raggiunto una classe ergodica, non la abbandona piú. Se una classe ergodica contiene un solo stato, tale stato si dice *assorbente*. Uno stato s_i è assorbente se, e solo se, $p_{ii} = 1$. Gli stati assorbenti si riconoscono facilmente perché ad ognuno di essi corrisponde, nella matrice di transizione, un elemento sulla diagonale principale eguale ad 1 (questo è, necessariamente l'unico elemento differente da zero della sua riga).

È importante studiare le classi ergodiche. Se s_i e s_j appartengono alla stessa classe ergodica, si introduca l'insieme dei *tempi di passaggio* da s_i a s_j

$$N_{ij} := \{n \in \mathbf{N} : p_{ij}^{(n)} > 0\} \quad (i, j = 1, 2, \dots, r).$$

In particolare, quando $s_i = s_j$ si parlerà di *tempi di ritorno* nello stato s_i; N_{ii} indicherà l'insieme dei tempi di ritono nello stato s_i.

Se n e s sono in N_{ii}, vi è anche $n + s$; infatti

$$p_{ii}^{(n+s)} = \sum_{h=1}^{r} p_{ih}^{(n)}\, p_{hi}^{(s)} \geq p_{ii}^{(n)}\, p_{ii}^{(s)} > 0.$$

N_{ii} è dunque stabile rispetto all'addizione; inoltre non è vuoto, per la definizione di classe.

È importante il seguente teorema della teoria elementare dei numeri.

Teorema 1.2.1. *Un insieme di numeri positivi che sia stabile rispetto all'addizione è formato da tutti i multipli del proprio massimo comun divisore, tranne, al piú, un numero finito di essi.*

Dimostrazione. Sia $J \subset \mathbf{N}$ l'insieme dell'ipotesi. Senza perdita di generalità, si può supporre che il massimo comun divisore (m.c.d.) di J sia 1. Se, infatti, il m.c.d. fosse $d > 1$, basterebbe dividere tutti gli elementi di J per d per ricondursi al caso $d = 1$. È noto, dall'algoritmo di EUCLIDE, che esisterà un insieme finito $\{n_1, n_2, \dots, n_k\} \subset J$ con m.c.d. eguale a 1 e che esistono numeri interi relativi $a_1, a_2, \dots, a_k \in \mathbf{Z}$ tali che $\sum_{i=1}^k a_i n_i = 1$. Si indichi con m la somma dei termini positivi e con $-n$ quella dei termini negativi dell'ultima somma:

$$m := \sum_{a_i>0} a_i n_i \quad \text{e} \quad -n := \sum_{a_i<0} a_i n_i;$$

perciò $m - n = 1$. Poiché J è stabile rispetto all'addizione, è $m, n \in J$. Sia ora $q \geq n(n-1)$; si può scrivere $q = an + b$, ove $a, b \in \mathbf{Z}_+$ con $a \geq n-1$ e $0 \leq b \leq n-1$. Perciò $q = an + b(m-n) = (a-b)n + bm \in J$. □

Sia d_i il m.c.d. di N_{ii}; per il teorema precedente, tutti gli elementi di N_{ii} sono multipli di d_i. Se s_i e s_j appartengono alla stessa classe di equivalenza risulta $N_{ij} \neq \emptyset$.

Teorema 1.2.2. *Se s_i e s_j appartengono alla stessa classe d'equivalenza si ha $d_i = d_j = d$.*

Dimostrazione. Si supponga che $a, b \in N_{ij}$ e che $c \in N_{ji}$; allora $a + c \in N_{ii}$, ma anche $a + kd_j + c$ è in N_{ii} se k è abbastanza grande. Poiché tanto $a + kd_j + c$ quanto $a + c$ sono divisibili per d_i, tale è anche la loro differenza kd_j e, siccome questo è vero per ogni k abbastanza grande, d_j è divisibile per d_i. In maniera analoga, si dimostra che d_i è multiplo di d_j. Perciò $d_i = d_j = d$. □

Il numero d che è caratteristico degli stati di ogni classe ergodica si dice *periodo* di quegli stati e di quella classe.

Se una catena non ha stati transienti e se vi è piú di una classe ergodica si tratta di due (o piú) catene di MARKOV che non interagiscono e che possono essere studiate separatamente. Si può, perciò, supporre, senza perdita di generalità, che l'intera catena sia composta da un'unica classe ergodica. Una tale catena si dirà *ergodica* o *irriducibile.*

1. Catene senza stati transienti.

1-A. L'unica classe ergodica è regolare: una tale catena si dice *regolare.*

1-B. L'unica classe ergodica è ciclica: una tale catena si dice *ciclica.*

2. Catene con stati transienti.

2-A. Tutte le classi ergodiche contengono un solo stato, che è necessariamente assorbente: una tale catena si dice *assorbente.*

2-B. Tutte le classi ergodiche sono regolari ma almeno una di esse contiene piú di uno stato.

2-C. Tutte le classi ergodiche sono cicliche.

2-D. La catena ha classi ergodiche sia cicliche sia regolari.

Cambiando in maniera opportuna il nome degli stati, è possibile porre la matrice di transizione di una catena di Markov in una forma che ne rende immediata la classificazione; se $U_1, U_2, \ldots, U_h$ sono le classi d'equivalenza, tale forma, detta *canonica*, è

$$\left|\begin{array}{lllllll}
P_1 & \vdots & & & & & \\
\cdots & \cdots & \cdots & & & & \\
R_2 & \vdots & P_2 & \vdots & & & \\
\cdots & \cdots & \cdots & & & & \\
R_3 & \vdots & & \vdots & P_3 & & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
 & \vdots & R_h & \vdots & & \vdots & P_h
\end{array}\right| .$$

P_i rappresenta la matrice di transizione, che può non essere una matrice stocastica, all'interno della classe U_i, mentre R_i regola il passaggio da stati della classe U_i a stati di altre classi, se ciò è possibile. R_i avrà gli elementi tutti nulli se, e solo se, U_i è una classe ergodica.

Si noti che, ai soli fini della classificazione delle catene di Markov, non è affatto necessario conoscere l'esatto valore degli elementi della matrice di transizione, quanto piuttosto sapere quali tra essi siano nulli e quali siano differenti da zero. Per esempio, per classificare la catena con matrice di transizione data da quella dell'esmpio 1.1.1. la matrice potrebbe essere scritta nella forma

$$\left|\begin{array}{ccccc}
x & 0 & 0 & 0 & 0 \\
x & 0 & x & 0 & 0 \\
0 & x & 0 & x & 0 \\
0 & 0 & x & 0 & x \\
0 & 0 & 0 & 0 & x
\end{array}\right| .$$

In questo modo si è indicata la presenza di un elemento diverso da zero scrivendo x al suo posto; è però evidente che con questa notazione non si vuole affatto indicare che x abbia lo stesso valore in ogni posizione. Nell'ultima matrice data, x ha, necessariamente, il valore 1 nella prima e nell'ultima riga, mentre non può essere eguale a 1 nelle altre righe.

Una catena di Markov può essere rappresentata schematicamente, ai soli fini della sua classificazione, in un altro modo: mediante un grafo orientato nel quale i vertici corrispondono agli stati e gli spigoli orientati ad una probabilità di transizione strettamente positiva. Per esempio, l'ultima matrice di transizione corrisponde alla rappresentazione data nella Figura 2.1. Lo studente dovrebbe abituarsi a fare uso dei diversi metodi nel far pratica di classificazione di catene di Markov.

Se una catena di Markov ha stati transienti, vi sono, essenzialmente, due problemi da risolvere: il comportamento della catena prima dell'ingresso in una classe ergodica, e quello dopo tale ingresso. Per quest'ultimo problema, lo studio è analogo a quello di una catena ergodica. Per il primo, ci si può, in ogni caso, ricondurre al comportamento di una catena assorbente, raggruppando cosí tutti gli stati di ogni classe ergodica in un unico stato assorbente fittizio; la nuova catena avrà cosí tanti stati ergodici quanti erano le classi ergodiche della catena originaria.

Fig. 2.1

Teorema 1.2.3. *In ogni catena di* Markov *finita, quale che sia lo stato iniziale, tende a* 1*, al tendere di* n *a* $+\infty$*, la probabilità* $p^{(n)}$ *che, al tempo* $t = n$*, il processo sia in uno stato ergodico. Se* s_i *e* s_j *sono stati transienti, esistono due costanti* $\beta > 0$ *e* $\gamma \in]0, 1[$ *tali che* $p_{ij}^{(n)} \leq \beta\gamma^n$ *per ogni* n *abbastanza grande.*

Dimostrazione. Si indichino, ora e nel seguito, con $\hat{T}$ e $\hat{E}$ le famiglie degli stati transienti ed ergodici, rispettivamente, della catena di Markov. Non vi è nulla da dimostrare se la catena è inizialmente in uno stato ergodico. Supporremo, pertanto, che la catena parta da uno stato transiente; dunque,

$$\sum_{s_i \in \hat{T}} p_i^{(0)} = 1.$$

Ricorrendo al teorema delle probabilità totali, si ha

$$p^{(n)} = \sum_{s_i \in \hat{E}} \mathbb{P}(X_n = s_i) = \sum_{s_i \in \hat{T}} \sum_{s_j \in \hat{E}} \mathbb{P}(X_n = s_j, X_0 = s_i) = \sum_{s_i \in \hat{T}} p_i^{(0)} \sum_{s_j \in \hat{E}} p_{ij}^{(n)},$$

ove $p_i^{(0)}$ rappresenta la probabilità che, all'istante iniziale, la catena si trovi nello stato $s_i \in \hat{T}$ ($p_i^{(0)} = \mathbb{P}(X_0 = s_i)$) e la doppia somma nell'ultima espressione rappresenta la probabilità che la catena, trovandosi all'istante iniziale in un qualsiasi stato transiente, si trovi, al tempo $t = n$, in uno stato ergodico. Per ogni stato $s_i \in \hat{T}$, esiste $n_i \in \mathbf{N}$ tale che

$$1 - \sum_{s_j \in \hat{T}} p_{ij}^{(n_i)} = \sum_{s_j \in \hat{E}} p_{ij}^{(n_i)} > 0;$$

perciò

$$\sum_{s_j \in \hat{T}} p_{ij}^{(n_i)} < 1.$$

Si noti che la successione $k \mapsto \sum_{s_j \in \hat{T}} p_{ij}^{(k)}$ è decrescente; infatti,

$$\sum_{s_j \in \hat{T}} p_{ij}^{(k+1)} = \sum_{s_j \in \hat{T}} \sum_{s_h \in \hat{T}} p_{ih}^{(k)} p_{hj} = \sum_{s_h \in \hat{T}} p_{ih}^{(k)} \sum_{s_j \in \hat{T}} p_{hj} \leq \sum_{s_h \in \hat{T}} p_{ih}^{(k)}.$$

Esiste, dunque, $c_i \in]0, 1[$ tale che, per ogni $k \geq n_i$, si abbia

$$\sum_{s_j \in \hat{T}} p_{ij}^{(k)} < c_i < 1;$$

poiché gli stati sono in numero finito, se si pone

$$c := \max\{c_i : s_i \in \hat{T}\}, \qquad \text{e} \qquad n_0 := \max\{n_i : s_i \in \hat{T}\},$$

si ha $c \in]0, 1[$ e, per ogni stato transiente $s_i \in \hat{T}$ e per ogni $k \geq n_0$,

$$\sum_{s_j \in \hat{T}} p_{ij}^{(k)} < c < 1.$$

Quindi, per ogni $s_i \in \hat{T}$, $k \geq n_0$ e $m \in \mathbf{Z}_+$,

$$\sum_{s_j \in \hat{T}} p_{ij}^{(mk+k)} = \sum_{s_h \in \hat{T}} p_{ih}^{(mk)} \sum_{s_j \in \hat{T}} p_{hj}^{(k)} \leq c \sum_{s_h \in \hat{T}} p_{ih}^{(mk)},$$

onde, per induzione,

$$\sum_{s_j \in \hat{T}} p_{ij}^{[(m+1)k]} \leq c \sum_{s_j \in \hat{T}} p_{ij}^{(mk)} \leq c^{m+1} \xrightarrow[m \to \infty]{} 0.$$

Una successione estratta da $\left\{\sum_{s_j \in \hat{T}} p_{ij}^{(k)} : k \in \mathbf{N}\right\}$ converge; e, poiché la successione è decrescente, l'intera successione converge allo stesso limite. Inoltre, per n con $m\,k < n < (m+1)\,k$, risulta

$$\sum_{s_j \in \hat{T}} p_{ij}^{(n)} \leq \sum_{s_j \in \hat{T}} p_{ij}^{(mk)} \leq c^m = \frac{1}{c}\, c^{m+1} < \frac{1}{c}\, c^{n/k}.$$

Basta ora prendere $\beta := 1/c$ e $\gamma := c^{1/k}$. □

Si dice che un vettore (riga) di probabilità α è *stazionario* per una catena di Markov con matrice di transizione Π, se accade che sia $\alpha\,\Pi = \alpha$. Dimostreremo che ogni catena di Markov ammette almeno un vettore di probabilità stazionario.

Teorema 1.2.4. *Per ogni matrice di transizione* Π *esiste un vettore di probabilità stazionario*

La dimostrazione piú semplice di questo teorema ricorre al teorema di punto fisso di Markov–Kakutani, che quindi premettiamo.

Definizione 1.2.1. Sia K un insieme convesso di $\mathbf{R}^r$. Si dice che $T : K \to K$ è *affine*, se, per ogni scelta di x e di y in K e di t in $[0,1]$ riesce $T\,[tx + (1-t)y] = t\,Tx + (1-t)\,Ty$.

In $\mathbf{R}^r$ considereremo la norma $\|x\| := \max\{|x_1|, |x_2|, \dots, |x_r|\}$.

Teorema 1.2.5. (Markov–Kakutani). *Ogni trasformazione continua e affine* T *di un sottoinsieme convesso e compatto* K *di* $\mathbf{R}^r$ *in sé ha un punto fisso.*

Dimostrazione. Si scelga $z \in K$ e si ponga $x_n := \frac{1}{n} \sum_{j=0}^{n-1} T^j z$, ove si conviene che T^0 sia l'operatore identità. In virtú della convessità di K, x_n è in K, e, per la compattezza di K, esistono $x \in K$ e una successione $(x_{n(k)})_{k \in \mathbf{N}}$ estratta da (x_n) tale che $\left\|x_{n(k)} - x\right\| \to 0$. Vogliamo ora dimostrare che x è un punto fisso per T, cioè, che $Tx = x$. L'insieme K è compatto e, dunque, limitato, sicché esiste finito $\lambda := \sup\{\|y\| : y \in K\}$. Perciò,

$$\begin{aligned}
\left\|x_{n(k)} - Tx_{n(k)}\right\| &= \left\| \frac{1}{n(k)} \sum_{j=0}^{n(k)-1} T^j z - \frac{1}{n(k)} \sum_{j=0}^{n(k)-1} T^{j+1} z \right\| \\
&= \frac{1}{n(k)} \left\|z - T^{n(k)} z\right\| \leq \frac{2\lambda}{n(k)} \xrightarrow[k \to +\infty]{} 0,
\end{aligned}$$

onde

$$\|x - Tx\| \leq \left\|x - x_{n(k)}\right\| + \left\|x_{n(k)} - Tx_{n(k)}\right\| + \left\|Tx_{n(k)} - Tx\right\|,$$

da cui segue l'asserto. □

Dimostrazione del Teorema 1.2.4. L'insieme convesso e compatto K è definito da

$$K := \left\{ x \in \mathbf{R}^r : x_j \geq 0 \quad (j = 1, 2, \dots, r), \quad \sum_{j=1}^{r} x_j = 1 \right\}.$$

Si definisca $T : \mathbf{R}^r \to \mathbf{R}^r$ mediante $Tx := x\,\Pi$, cioè $(Tx)_j := \sum_{i=1}^{r} x_i p_{ij}$. La trasformazione cosí definita è continua; infatti, ponendo

$$c := \max_{j=1,2,\dots,r} \sum_{i=1}^{r} p_{ij}$$

si ha, poiché T è addirittura lineare, come trasformazione da $\mathbf{R}^r$ in $\mathbf{R}^r$,

$$\|Tx\| = \max_{j=1,2,\dots,r} \sum_{i=1}^{r} x_i\, p_{ij} \leq \|x\| \max_{j=1,2,\dots,r} \sum_{i=1}^{r} p_{ij} \leq c\,\|x\|.$$

Per poter applicare il Teorema di Kakutani basta mostrare che $TK \subset K$, vale a dire che T manda K in sé. Se $x \in K$, si ha $(Tx)_i \geq 0$ e

$$\sum_{j=1}^{r} (Tx)_j = \sum_{j=1}^{r} \sum_{i=1}^{r} x_i\, p_{ij} = \sum_{i=1}^{r} x_i \sum_{j=1}^{r} p_{ij} = \sum_{i=1}^{r} x_i = 1;$$

perciò $Tx \in K$.

1.3 Catene assorbenti

La forma canonica della matrice di transizione con $r - k$ stati assorbenti e k stati transienti è data da

$$\begin{vmatrix} I_{r-k} & O \\ R & Q \end{vmatrix}$$

ove I_{r-k} è la matrice identità $(r-k) \times (r-k)$, Q è una matrice $k \times k$ e O è la matrice di ordine $(r-k) \times k$ con tutti gli elementi nulli.

Teorema 1.3.1. *Per ogni catena assorbente, la matrice $I_k - Q$ ammette inversa e risulta*

$$(I_k - Q)^{-1} = \sum_{n \in \mathbf{Z}_+} Q^n.$$

Dimostrazione. Vale l'identità

$$(I_k - Q)\,(I_k + Q + Q^2 + \cdots + Q^{n-1}) = I_k - Q^n.$$

È evidente dal Teorema 1.2.3 e dalla forma canonica, che $Q^n \to O_k$, ove O_k è la matrice con tutti gli elementi nulli di ordine $k \times k$. Allora, per n sufficientemente grande, si ha $\det(I_k - Q^n) \neq 0$ e perciò $\det(I_k - Q) \neq 0$. □

Definizione 1.3.1. Si dice *matrice fondamentale* di una catena assorbente la matrice $A := (I_k - Q)^{-1}$.

Si noti che

$$A := (I_k - Q)^{-1} = \sum_{n \in \mathbf{Z}_+} Q^n.$$

Il significato della matrice fondamentale è dato dal seguente.

Teorema 1.3.2. *Per una matrice assorbente sia $\hat{T}$ l'insieme degli stati transienti e, per $s_j \in \hat{T}$, si indichi con N_j la v.a. che conta il numero di istanti, compreso, eventualmente, quello iniziale, nei quali il processo è nello stato s_j, cioè $N_j = \sum_{n \in \mathbf{Z}_+} 1_{\{X_n = s_j\}}$. Allora*

$$\{E_i(N_j)\} = A \qquad (s_i, s_j \in \hat{T}),$$

ove E_i indica la speranza calcolata usando le probabilità condizionate p_{ih} $(h = 1, 2, \dots, r)$.

Dimostrazione. Poiché $N_j = \sum_{n \in \mathbf{Z}_+} 1_{\{X_n = s_j\}}$,

$$\begin{aligned} \{E_i(N_j)\} &= \left\{ E_i \left(\sum_{n \in \mathbf{Z}_+} 1_{\{X_n = s_j\}} \right) \right\} \\ &= \sum_{n \in \mathbf{Z}_+} \{p_{ij}^{(n)}\} = \sum_{n \in \mathbf{Z}_+} Q^n = A, \end{aligned}$$

□

che conclude la dimostrazione.

Teorema 1.3.3. *Se s_i è uno stato transiente, $s_i \in \hat{T}$, e s_j è assorbente, $s_j \in \hat{E}$, e se b_{ij} è la probabilità che il processo parta da s_i e termini in s_j, risulta*

$$B := \{b_{ij}\} = AR.$$

Dimostrazione. Si osservi che la catena, partendo dallo stato s_i, può andare direttamente, e in un passo, da s_i a s_j, oppure può andare da s_i ad un altro stato transiente s_h e di lí poi a s_j. Vale, dunque, la relazione

$$b_{ij} = p_{ij} + \sum_{s_k \in \hat{T}} p_{ik}\, b_k, \tag{1.3.1}$$

equazione che si scrive, in forma matriciale, $B = R + Q\,B$, onde $B = A\,R$. □

1.4 Catene regolari

Sono queste le catene di Markov che abbiamo classificato come del tipo **1-A**, cioè senza stati transienti e con un'unica classe ergodica che contiene una sola classe ciclica. Si dirà *regolare* la matrice di transizione di una catena di Markov regolare. Il teorema che segue è ovvia conseguenza delle definizioni. Notiamo dapprima che, se una matrice di transizione Π ha una potenza, sia Π^n, con tutti gli elementi diversi da zero, allora tali sono anche tutti gli elementi di Π^{n+1}. Infatti, fissato lo stato s_i, esiste sicuramente almeno uno stato s_h con $p_{ih} > 0$; allora, quale che sia lo stato s_k, si ha

$$p_{ik}^{(n+1)} \geq p_{ih}\, p_{hk}^{(n)} > 0.$$

Teorema 1.4.1. *Una matrice di transizione* Π *è regolare se, e solo se, esiste un numero naturale* n *tale che* Π^n *abbia tutti gli elementi positivi.*

Dimostrazione. Si supponga che la matrice di transizione Π sia regolare e siano s_i e s_j stati qualsiasi; esiste allora $k \in \mathbf{N}$ tale che $p_{ij}^{(k)} > 0$. Poiché $d = 1$, per h sufficientemente grande, si ha

$$p_{ij}^{(k+h)} \geq p_{ij}^{(k)} \, p_{jj}^{(h)} > 0.$$

Dunque, per ogni coppia di stati s_i e s_j esiste $n(i,j) \in \mathbf{N}$ tale che

$$\Pi^{n(i,j)} > 0.$$

Alla luce dell'osservazione fatta sopra, basta ora prendere il minimo comune multiplo n dei numeri $n(i,j)$, che sono, al piú, r^2, per avere $\Pi^n > 0$.

Viceversa, si pupponga che esista $n \in \mathbf{N}$ tale che Π^n abbia tutti gli elementi diversi da zero. Si è visto sopra che, in tal caso, anche tutte le potenze di Π con esponente maggiore di n hanno gli elementi tutti diversi da zero. È perciò finito l'insieme dei naturali n per i quali esiste una coppia di indici i e j tali che $p_{ij}^{(n)} = 0$. Poiché tra i numeri $n \in \mathbf{N}$ per i quali $p_{ii}^{(n)} > 0$ ve ne sono certamente due primi tra loro, si ha $d = 1$. Ogni catena di Markov con matrice di transizione Π è dunque regolare. □

Lemma 1.4.1. *Sia* Π *una matrice di transizione* $r \times r$ *con tutti gli elementi strettamente positivi e sia* $\varepsilon > 0$ *il minimo elemento di* Π. *Sia* x *un qualsiasi vettore colonna a* r *componenti* $x = (x_1, x_2, \dots, x_r)^T$*; posto* $m_0 := \min\{x_1, x_2, \dots, x_r\}$ *e* $M_0 := \max\{x_1, x_2, \dots, x_r\}$*, siano* m_1 *e* M_1 *le analoghe quantità per* Πx. *Allora, è* $M_1 \leq M_0$*,* $m_1 \geq m_0$ *e*

$$M_1 - m_1 \leq (1 - 2\varepsilon)(M_0 - m_0). \tag{1.4.1}$$

Si osservi che la (1.4.1) ha significato anche quando sia

$$\Pi = \begin{vmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{vmatrix};$$

in questo caso, e solo in questo, si ha infatti $\varepsilon = \frac{1}{2}$. Nel seguito escluderemo questo caso banale e supporremo che sia $\varepsilon < \frac{1}{2}$.

Dimostrazione. Sia x' un vettore ottenuto da x ponendo tutte le componenti eguali a M_0, tranne una di quelle che sono eguali m_0; si ha, cosí, $x \leq x'$. Ogni componente del vettore colonna $\Pi x'$ è della forma $a\, m_0 + (1-a)\, M_0 = M_0 - a\,(M_0 - m_0)$, con $a \geq \varepsilon$ e perciò risulta minore di, o, al piú eguale a, $M_0 - \varepsilon\,(M_0 - m_0)$. Ma, poiché è $x \leq x'$, si ha $M_1 \leq M_0 - \varepsilon(M_0 - m_0)$. Lo stesso ragionamento applicato a $-x$ dà $-m_1 \leq -m_0 - \varepsilon(-m_0 + M_0)$, onde, sommando, si ottiene la (1.4.1). □

Si noti che le diseguaglianze $M_1 \leq M_0$ e $m_1 \geq m_0$ valgono per ogni matrice stocastica.

Teorema 1.4.2. *Se* Π *è una matrice di transizione regolare, allora*

(a) $\lim_{n\to+\infty} \Pi^n = A$, *ove* A *è una matrice stocastica;*

(b) $A = \xi_r \alpha$, *ove* α *è il vettore riga* $(a_1, a_2, \ldots, a_r)$ *(le righe di* A *sono dunque tutte eguali);*

(c) $a_i > 0 \quad (i = 1, 2, \ldots, r)$.

Dimostrazione. Si supponga, dapprima, che tutti gli elementi di Π siano strettamente positivi. Sia ε il minimo elemento di Π e sia ρ_j il vettore colonna di componenti $\delta_{ij} \quad (i = 1, 2, \ldots, r)$. Siano M_n e m_n, rispettivamente, la massima e la minima componente di $\Pi^n \rho_j$. Poiché è $\Pi^n \rho_j = \Pi\, \Pi^{n-1} \rho_j$, il lemma 1.4.1 implica $M_1 \geq M_2 \geq \cdots \geq M_n$ e $m_1 \leq m_2 \cdots \leq m_n$; inoltre $M_n - m_n \leq (1 - 2\varepsilon)(M_{n-1} - m_{n-1})$, onde, iterando, e ponendo $\delta_n := M_n - m_n$, $\delta_n \leq (1 - 2\varepsilon)^n \, \delta_0$. Poiché $\delta_0 = 1$ e $\varepsilon < 1/2$, segue che $\delta_n \to 0$, ciò che significa che $\Pi^n \rho_j$ tende ad un vettore colonna avente tutte le componenti eguali. Sia a_j il loro valore comune. Poiché

$$0 < m_1 \leq \cdots \leq m_n \leq a_j \leq M_n \leq \cdots \leq M_1 < 1,$$

è $a_j \in \,]0, 1[$. Se si considera che $\Pi^n \rho_j$ è la j–esima colonna di Π^n, e che quest'ultima è una matrice stocastica, si vede che il teorema è completamente dimostrato, nell'ipotesi fatta.

Se, poi, Π è solo regolare, esiste un naturale N tale che Π^N abbia tutti elementi strettamente positivi. Sia $\varepsilon' > 0$ il piú piccolo tra questi. Procedendo come sopra si ha $\delta_{kN} < (1 - 2\varepsilon')^k$. Ora, la successione $\{\delta_n\}$ è decrescente e da questa si può estrarre un'altra successione $\{\delta_{kN}\}$ che tende a zero. Perciò tutta la successione tende a tale limite. La dimostrazione procede poi come nel caso precedente. □

Il Teorema 1.4.2 asserisce che $\lim_{n\to+\infty} p_{ij}^{(n)} = a_j$, quale che sia lo stato s_i. In particolare vale il seguente

Corollario 1.4.1. *Se* Π *è una matrice di transizione regolare e*

$$a_j = \lim_{n\to+\infty} p_{ij}^{(n)},$$

esistono due costanti $b > 0$ *e* $c \in \,]0, 1[$ *tali che*

$$p_{ij}^{(n)} = a_j + e_{ij}^{(n)}$$

con $|e_{ij}^{(n)}| \leq bc^n$.

Dimostrazione. Segue dal Teorema 1.4.2 che $|e_{ij}^{(n)}| \leq d_n$. Sia N il minimo naturale tale che Π^N non abbia elementi nulli. Si scelgano $c := (1-2\varepsilon)^{1/N}$ e $b := (1-2\varepsilon)^{-1} = c^{-N}$. Se $n = kN$, allora $d_n \leq (1 - 2\varepsilon)^k = c^n$. Se, invece, $n = kN + n_1$ con $0 < n_1 < N$, allora, poiché $\{d_n\}$ è decrescente, $d_n \leq d_{n-n_1} \leq c^{n-n_1} \leq c^n c^{-N} = bc^n$. □

Teorema 1.4.3. *Se* Π *è una matrice di transizione regolare e* A *e* α *sono definiti come nel Teorema* 1.4.2*, allora:*

(a) $\lim_{n\to+\infty} p\Pi^n = \alpha$, *quale che sia il vettore riga di probabilità* p*;*

(b) $A\Pi = \Pi A = A$*;*

(c) α è l'unico vettore stazionario di Π (cioè tale che $\alpha\Pi = \alpha$); inoltre α è un vettore stazionario di A.

Dimostrazione. (a) Se p è un vettore riga di probabilità, si ha $p\xi_r = 1$, onde $pA = p\xi_r\alpha = \alpha$. Perciò $p\Pi^n \xrightarrow[n\to+\infty]{} pA = \alpha$.

(b) $\Pi^{n+1} = \Pi^n\Pi$ implica $A = A\Pi$; similmente per $A = \Pi A$.

(c) $\alpha\Pi = \alpha$ segue dallo scrivere le righe della relazione matriciale $A\Pi = A$. Per dimostrare l'unicità di α, si supponga che, per un altro vettore riga di probabilità β, sia $\beta\Pi = \beta$. Di qui, e dalla (1), segue $\beta = \beta\Pi = \beta\Pi^2 = \cdots = \beta\Pi^n \xrightarrow[n\to+\infty]{} \alpha$, sicché $\beta = \alpha$. □

1.5 Catene cicliche

I risultati della sezione precedente possono essere estesi alle catene cicliche, per le quali, come sappiamo, è $d > 1$. Alla luce del Teorema 1.4.1 si può dire che nessuna potenza di Π avrà componenti tutte positive. La posizione degli zeri cambierà ciclicamente al cambiare dell'esponente. Ciò vuol dire che la successione (Π^n) non converge, almeno nel senso usuale. È questa la differenza principale tra una catena ciclica ed una regolare. Si vedrà tra breve che vale però un risultato piú debole.

Definizione 1.5.1. Si dice che una successione (x_n) di numeri reali *converge nel senso di* EULERO se esiste $s \in]0,1[$ tale che sia convergente la successione

$$\left\{ \sum_{j=0}^{n} \binom{n}{j} s^{n-j}(1-s)^j x_j : n \in \mathbf{Z}_+ \right\}.$$

Per la dimostrazione del seguente teorema rimandiamo agli eserciziî.

Teorema 1.5.1. *Se la successione (x_n) converge nel senso usuale a $x \in \mathbf{R}$, vi converge anche nel senso di* EULERO *per ogni $s \in]0,1[$.*

Il seguente è l'analogo del Teorema 1.4.2.

Teorema 1.5.2. *Se Π è la matrice di transizione di una catena ciclica, la successione (Π^n) converge nel senso di* EULERO *ad una matrice $A = \xi_r\,\alpha$ ove $\alpha = (a_1, a_2, \dots, a_r)$ ha solo componenti strettamente positive.*

Dimostrazione. Per $s \in]0,1[$, è facile controllare che $s\,I_r + (1-s)\,\Pi$ è una matrice stocastica. Inoltre nella posizione nella quale Π ha un elemento strettamente positivo, anche la nuova matrice ha un elemento strettamente positivo, sicché quest'ultima rappresenta una catena ergodica. Poiché gli elementi diagonali sono maggiori di $s > 0$, è possibile ritornare, in un passo, ad ogni stato; perciò $d = 1$ e la nuova catena è regolare. Il Teorema 1.4.2 assicura ora che la successione $([s\,I_r + (1-s)\,\Pi]^n)$ tende a $A = \xi_r\,\alpha$ con $\alpha > 0$. Cosí

$$A = \lim_{n\to+\infty} \{s\,I_r + (1-s)\,\Pi\}^n = \lim_{n\to+\infty} \sum_{j=0}^{n} \binom{n}{j} s^{n-j}(1-s)^j \Pi^j, \tag{1.5.1}$$

che fornisce l'asserto. □

Si osservi che (Π^n) converge nel senso di EULERO per ogni valore s di $]0,1[$.

Teorema 1.5.3. *Sia Π una matrice di transizione ciclica e sia A la matrice del teorema precedente. Allora:*

(a) per ogni vettore (riga) di probabilità p, $(p\,\Pi^n)$ converge a α nel senso di EULERO*;*

(b) α è l'unico vettore di probabilità stazionario per Π, $\alpha\,\Pi = \alpha$;

(c) $A\,\Pi = \Pi\,A = A$.

Dimostrazione. (a) Basta moltiplicare la (1.5.1), a sinistra, per p per vedere che $(p\Pi^n)$ converge nel senso di EULERO a $p\,A = p\xi_r\,\alpha = \alpha$.

(b) Come nel Teorema 1.4.3(c), α è l'unico vettore di probabilità stazionario per $s\,I_r + (1-s)\,\Pi$, con $s \in\,]0,1[$. Se si mostra che le matrici di transizione Π e $s\,I_r + (1-s)\,\Pi$ hanno gli stessi vettori stazionarî, si sarà anche mostrato l'asserto. Ora $\beta\,\{s\,I_r + (1-s)\,\Pi\} = \beta$ se, e solo se, $(1-s)\beta\,\Pi = (1-s)\beta$, ossia, poiché $1-s \neq 0$, $\beta\,\Pi = \beta$.

(c) Per ogni matrice stocastica Π si ha $\Pi\,\xi_r = \xi_r$, sicché dalla (b) scendono

$$\Pi\,A = \Pi\,(\xi_r\,\alpha) = (\Pi\,\xi_r)\,\alpha = \xi_r\,\alpha = A$$

e

$$A\,\Pi = (\xi_r\,\alpha)\Pi = \xi_r\,(\alpha\,\Pi) = \xi_r\alpha = A,$$

che conclude la dimostrazione. □

1.6 Due complementi

Diamo infine due complementi sulle probabilità limite e su quelle stazionarie.

Teorema 1.6.1. *Sono equivalenti le due proprietà:*

(a) esistono i limiti $\lim_{n\to+\infty} p_{ij}^{(n)} = a_j \quad (i = 1, 2, \dots, r)$;

(b) esiste $n \in \mathbf{N}$ tale che Π^n abbia una colonna di elementi strettamente positivi.

Dimostrazione. L'implicazione (b) $\Longrightarrow$ (a) si dimostra come nel Teorema 1.4.2(a).

(a) $\Longrightarrow$ (b) Poiché $\sum_{j=1}^n a_j = 1$, almeno una delle componenti di α, per esempio a_k, è strettamente positiva. Da $\lim_{n\to+\infty} p_{ik}^{(n)} = a_k$, scende che esiste $n_i \in \mathbf{N}$ tale che $p_{ik}^{(n)} > 0$ non appena $n \geq n_i$. Posto $n_0 := \max\{n_i : i = 1, 2, \dots, r\}$ risulta, se $n \geq n_0$, $p_{ik}^{(n)} > 0$ per ogni $i \leq r$, cioè la k–colonna di Π^n è composta di elementi strettamente positivi. □

Teorema 1.6.2. *Ogni catena di* MARKOV *con una sola classe ergodica (che si dice* indecomponibile*) ammette un solo vettore di probabilità stazionario. Questo vettore ha componenti strettamente positive in corrispondenza degli stati ergodici, nulle in corrispondenza di quelli transienti.*

Dimostrazione. La matrice di transizione scritta in forma canonica è

$$\Pi = \begin{bmatrix} S & 0 \\ R & Q \end{bmatrix}$$

ove S è la matrice di transizione della classe ergodica (che da sola forma una catena ergodica). Allora, tanto che S sia regolare quanto che sia ciclica, S ammette un unico vettore di probabilità stazionario $\alpha_1 > 0$: $\alpha_1 S = \alpha_1$. Si ponga $\alpha := (\alpha_1, \alpha_2)$ con $\alpha_2 := (0, 0, \ldots, 0)$. È immediato verificare che $\alpha \Pi = \alpha$. Per dimostrare l'unicità di α, basta verificare che è unico α_2, poiché già si sa che è unico il vettore stazionario α_1. Se $\beta = (\alpha_1, \beta_2)$ è un vettore di probabilità stazionario, risulta necessariamente $\beta_2 = \alpha_2 = 0$, perché la somma delle componenti di α_1 è uno e anche quella delle componenti di β è uno. $\square$