Appunti di Probabilità 2^a U.D.

Luciano Tubaro

anno accademico 2003-2004

Indice

Riprendiamo il filo del discorso dal “concetto” di variabile casuale. Abbiamo chiamato variabile casuale una “grandezza” (fisica) che si “comporta” in modo “aleatorio”, quindi il cui valore non è determinato a priori. la precedente non è una definizione matematica di variabile casuale, è più un *modus operandi.*

Consideriamo un esempio concreto: sia T il tempo di durata di una data lampadina. Qui T, la durata, è una grandezza ben individuata. Prima di accendere la lampadina e anche durante il funzionamento non è possibile affermare una durata precisa. Certo ci sarà un momento in cui la lampadina smetterà di funzionare, diciamo t_R: potremo solo allora affermare che $T = t_R$. Quando introduciamo la locuzione “comportamento aleatorio” pensiamo che la lampadina avrebbe potuto smettere di funzionare anche in un tempo t'_R, nel senso che se potessimo ripetere l'esperimento non ci aspetteremmo di nuovo che $T = t_R$, ma un valore diverso. Ma al posto di affermare che T sarà uguale ad un valore t pensiamo di poter parlare di una famiglia di numeri (tra 0 ed 1)

$$\Pr(a < T \leqslant b)$$

al variare di a e b, ciascuno dei quali rappresenta il grado di fiducia che T sia uguale ad un valore tra a e b. Se consideriamo solo i numeri $F(t) = \mathbb{P}(T \leqslant t)$ al variare di t, recuperiamo gli altri dalla formula

$$\Pr(a < T \leqslant b) = F(b) - F(a)$$

Riassumiamo: prima di accendere la lampadina, **invece** di prevedere il valore di T, cioè affermare che $T = t_0$, cioè scegliamo un determinato t_0, consideriamo una funzione $F(t)$. Quindi, apparentemente, passiamo da una affermazione più semplice $T = t_0$ ad una affermazione (apparentemente) molto più complessa e nello stesso tempo meno precisa, cioè la scelta di $F(t)$ e lavorare con questa funzione.

Notiamo ora che gli eventi che c'interessano, $(a < T \leqslant b)$, posso metterli in corrispondenza con gli intervalli $]a, b]$ di $\mathbb{R}$. Ora, utilizzando un risultato noto della teoria della misura, dalla definizione

$$\mu(]a, b]) = F(b) - F(a)$$

posso estendere la μ a tutta la famiglia $\mathscr{B}$ dei boreliani di $\mathbb{R}$. μ risulta una misura (di probabilità) sui boreliani; abbiamo definito in altre parole una terna $(\mathbb{R}, \mathscr{B}, \mu)$. Una scelta diversa di F, diciamo F', avrebbe dato luogo ad una corrispondente misura μ', diversa dalla precedente μ.

1 Formalizzazione

Sia dato uno **spazio di probabilità** $(\Omega, \mathscr{E}, \mathbb{P})$, cioè una terna dove

- Ω è un insieme (da scegliere a seconda del problema)
- $\mathscr{E}$ è una σ-algebra di sottoinsiemi di Ω (da scegliere a seconda del problema); si ha evidentemente

$$\{\emptyset, \Omega\} \subset \mathscr{E} \subset \mathscr{P}(\Omega) \ \ (= 2^\Omega)$$

- $\mathbb{P}$ è una misura (di probabilità) σ-additiva (numerabilmente additiva) su $\mathscr{E}$ (da scegliere a seconda del problema)

Definizione 1 *Si dice* **variabile casuale** *una funzione* $X\colon \Omega \to \mathbb{R}$ *che sia* $\mathscr{E}$*-misurabile, cioè se per ogni insieme boreliano* $I \subset \mathbb{R}$ *l'immagine inversa* $X^{-1}(I)$ *è un elemento di* $\mathscr{E}$*.*

Indicando con $\mathscr{B}$ la σ-algebra dei boreliani di $\mathbb{R}$, è facile convincersi che la famiglia $\mathscr{F}_X$ di sottoinsiemi di Ω definita da

$$\mathscr{F}_X = \{X^{-1}(I), I \in \mathscr{B}\}$$

è una σ-algebra in Ω. Dire allora che X è una variabile casuale equivale a dire che $\mathscr{F}_X \subset \mathscr{E}$. Potremo ulteriormente dire che $\mathscr{F}_X$ è la più piccola σ-algebra che rende X misurabile.

Data una variabile casuale X possiamo introdurre la misura associata

$$\mu_X(I) = \mathbb{P}(X^{-1}(I))$$

sui boreliani $I \in \mathscr{B}$ di $\mathbb{R}$. Questa misura μ_X si dice **distribuzione di probabilità** di X. Verificare che μ_X è una misura σ-additiva su $\mathbb{R}$. μ_X è la misura immagine di $\mathbb{P}$ tramite la variabile casuale X. Come abbiamo già detto in precedenza la misura μ_X risulta univocamente determinata solo dai suoi valori sui boreliani I del tipo $]-\infty, t]$, in altre parole risulta univocamente determinata dalla sua **funzione di ripartizione** F_X

$$F_X(t) = \mu_X(]-\infty, t]) = \mathbb{P}(X \le t)$$

1.1 Integrazione

Dato una spazio di probabilità $(\Omega, \mathscr{E}, \mathbb{P})$ ed una variabile casuale $X \geq 0$ si può definire sempre il suo integrale (rispetto a $\mathbb{P}$) che indicheremo con

$$\int_\Omega X(\omega)\mathbb{P}(d\omega)$$

o a volte più semplicemente con

$$\int_\Omega X \, \mathrm{d}\mathbb{P}$$

ma che può essere anche $+\infty$! Quando $\int_\Omega X \, \mathrm{d}\mathbb{P} < +\infty$ diremo che X è **integrabile**.

Più in generale data una variabile casuale X, possiamo scrivere (in modo univoco)

$$X(\omega) = X^+(\omega) - X^-(\omega)$$

dove $X^+ = \max\{X, 0\}$ e $X^- = \max\{-X, 0\} = (-X)^+$, che si dicono rispettivamente la parte positiva e la parte negativa di X. Inoltre risulta $|X| = X^+ + X^-$. Richiedere che $|X|$ sia integrabile equivale a richiedere che entrambi, X^+ e X^-, siano integrabili; in tal caso definiremo

$$\mathbb{E}(X) = \int_\Omega X \, \mathrm{d}\mathbb{P} = \int_\Omega X^+ \, \mathrm{d}\mathbb{P} - \int_\Omega X^- \, \mathrm{d}\mathbb{P}$$

che chiameremo **valore di aspettazione** ([1]).

Se il quadrato X^2 di X è integrabile, è integrabile anche il modulo $|X|$ e potremo definire la cosiddetta **varianza** di X:

$$\mathbf{Var}(X) = \mathbb{E}\Big((X - \mathbb{E}(X))^2\Big) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Vale la seguente fonddamentale formula (di cambiamento di variabili)

$$\mathbb{E}(f(X)) = \int_\Omega f(X) \, \mathrm{d}\mathbb{P} = \int_\mathbb{R} f(x) \, \mu_X(\mathrm{d}x) = \int_\mathbb{R} f(x) \, \mathrm{d}F_X(x) \qquad (1)$$

2 Funzioni caratteristiche

Un altro modo per caratterizzare la distribuzione di probabilità μ ([2]) della variabile casuale X è di introdurre la seguente funzione (a valori complessi $\mathbb{C}$)

$$\varphi(t) = \int_\mathbb{R} e^{itx}\mu(\mathrm{d}x) = \int_\mathbb{R} e^{itx} \, \mathrm{d}F = \mathbb{E}(e^{itX})$$

[1] Oppure ASPETTAZIONE, SPERANZA, SPERANZA MATEMATICA, VALOR MEDIO, MEDIA.
[2] Ometteremo il sottoindice X per semplicità e quando non genererà confusione.

che si chiama **funzione caratteristica** di μ (o anche funzione caratteristica della variabile casuale X).

Una funzione caratteristica è caratterizzata (teorema di Bochner) dalle seguenti proprietà

- $\varphi(0) = 1$,
- φ è continua in 0,
- per ogni scelta di n, di n valori $t_1, t_2, \ldots, t_n$ e di n numeri complessi $z_1, z_2, \ldots, z_n$ risulta
$$\sum_{i,j=1}^{n} \varphi(t_i - t_j) z_i \bar{z}_j \geqslant 0.$$

La terza proprietà si può anche esprimere dicendo che la funzione φ è definita positiva.

Se φ è una funzione caratteristica allora anche $\bar{\varphi}$ è una funzione caratteristica. Una combinazione convessa di funzioni caratteristiche è ancora una funzione caratteristica e la parte reale $\mathrm{Re}\,\varphi$ è una funzione caratteristica. Il prodotto di funzioni caratteristiche è ancora una funzione caratteristica e il modulo al quadrato è ancora una funzione caratteristica.

3 Varie

Data una successione di eventi $A_1, A_2, A_3, \ldots$, si definisce $\limsup A_k$ l'evento

$$\limsup A_k = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

Lemma 1 (Borel-Cantelli) *Sia $A_1, A_2, A_3, \ldots$ una successione di eventi, tali che*

$$\sum_{k=1}^{\infty} \mathbb{P}(A_k) < \infty,$$

allora l'evento $A = \limsup A_k$ ha probabilità nulla, $\mathbb{P}(A) = 0$. Naturalmente si ha che $\mathbb{P}(A^c) = 1$ e $A^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c$.

Dimostrazione: Dalla definizione abbiamo che $A \subset \bigcup_{k=n}^{\infty} A_k$. Quindi

$$\mathbb{P}(A) \leq \sum_{k=n}^{\infty} \mathbb{P}(A_k)$$

che per ipotesi tende a zero quando $n \to \infty$. ∎

4 Catene di Markov

Sia $\mathscr{S} = \{1, 2, 3, \ldots, N\}$, che chiameremo *stati*, e consideriamo una successione di variabili casuali $\{X_n, n \geq 0\}$ a valori in $\mathscr{S}$.

Diremo che la successione $\{X_n, n \geq 0\}$ è una **catena di Markov** se risulta verificata la seguente proprietà

$$\mathbb{P}(X_{k+n} = j | X_k = i, X_{k-1} = i_{k-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{k+n} = j | X_k = i) \quad (2)$$

Diremo che la catena di Markov $\{X_n, n \geq 0\}$ è **stazionaria** se, per ogni k, risulta

$$\mathbb{P}(X_{k+n} = j | X_k = i) = \mathbb{P}(X_n = j | X_0 = i) =: p_{ij}^{(n)} \quad (3)$$

Anche per le catene di Markov ([3]) (che sono particolari **processi di Markov**) vale la relazione di Chapman-Kolmogorov, che assume la forma seguente

$$p_{ij}^{(n+m)} = \sum_{k=1}^{N} p_{ik}^{(n)} p_{kj}^{(m)} \quad (4)$$

Nota: In particolare dalla (4) *discende*

$$p_{ij}^{(n+m)} \geq p_{ik}^{(n)} p_{kj}^{(m)} \quad (5)$$

Ancora, sempre per la (4)*, avremo che* $p_{ij}^{(n+m+s)} = \sum_{k=1}^{N} p_{ik}^{(n+m)} p_{kj}^{(s)} = \sum_{k=1}^{N} \sum_{l=1}^{N} p_{il}^{(n)} p_{lk}^{(m)} p_{kj}^{(s)}$ *da cui*

$$p_{ij}^{(n+m+s)} \geq p_{il}^{(n)} p_{lk}^{(m)} p_{kj}^{(s)} \qquad \forall l, \forall k \quad (6)$$

Denotiamo con $p_{ij} = p_{ij}^{(1)}$ (osserviamo che $p_{ij}^{(0)} = \delta_{ij}$, dove $\delta_{ij} = 0$ se $i \neq j$ e $\delta_{ii} = 1$); è facile convincersi allora che la matrice formata dagli elementi $p_{ij}^{(n)}$ è esattamente P^n, dove P è la matrice formata dagli elementi p_{ij} (semplice applicazione dell'equazione (4)). Risulta inoltre facilmente che

$$\sum_{j=1}^{N} p_{ij}^{(n)} = 1$$

per ogni n.

L'obiettivo finale è indagare il comportamento di P^n per $n \to \infty$ (comportamento asintotico) ovvero studiare

$$\lim_{n \to \infty} p_{ij}^{(n)}$$

L'esistenza o no del suddetto limite dipende dagli stati i e j; introdurremo una classificazione degli stati in $\mathscr{S}$ in modo da poter essere in grado di enunciare il teorema riguardo al comportamento asintotico.

[3]da ora in poi SOTTINTENDEREMO l'aggettivo "stazionaria".

4.1 Classificazione degli stati

Definizione 2 *Diremo che due stati i e j di $\mathscr{S}$ sono* intercomunicanti, *se esistono due interi non negativi n e m tali che*

$$p_{ij}^{(n)} > 0 \qquad e \qquad p_{ji}^{(m)} > 0 \tag{7}$$

Osservazione. È facile verificare che la relazione $i \sim j$, dove con $\sim$ indichiamo che i e j sono intercomunicanti, è una relazione di equivalenza. Verifichiamo per esempio la transitività: $i \sim j$, $j \sim k \Rightarrow i \sim k$; ricordiamo che $i \sim j$ significa che esistono due interi non negativi n_1, n_2 tali che $p_{ij}^{(n_1)} > 0$, $p_{ji}^{(n_2)} > 0$; analogamente per $j \sim k$ esistono altri due interi non negativi m_1, m_2 tali che $p_{jk}^{(m_1)} > 0$, $p_{kj}^{(m_2)} > 0$; basta allora osservare che, per la (5),

$$p_{ik}^{(n_1+m_1)} \geqslant p_{ij}^{(n_1)} p_{jk}^{(m_1)} > 0, \qquad p_{ki}^{(n_2+m_2)} \geqslant p_{kj}^{(m_2)} p_{ji}^{(n_2)} > 0$$

cioè $i \sim k$. Lasciamo come esercizio provare la riflessività e la simmetria. ∎

Come conseguenza l'insieme degli stati $\mathscr{S}$ si può dividere in classi disgiunte di elementi equivalenti $\mathscr{C}_1, \mathscr{C}_2, \ldots, \mathscr{C}_M$.

Definizione 3 *Diremo che un sottoinsieme $\mathscr{K}$ di stati ($\mathscr{K} \subset \mathscr{S}$) è* chiuso *se, per ogni $i \in \mathscr{K}$ e per ogni $j \notin \mathscr{K}$, risulta $p_{ij} = 0$.*

Proposizione 1 *Se $\mathscr{K}$ è un sottoinsieme chiuso e $\mathscr{C}_\nu$ una qualunque classe di equivalenza allora*

$$\mathscr{K} \cap \mathscr{C}_\nu = \emptyset \qquad \textit{oppure} \qquad \mathscr{C}_\nu \subset \mathscr{K}$$

Dimostrazione: Supponiamo $\mathscr{K} \cap \mathscr{C}_\nu \neq \emptyset$; osserviamo che se $\mathscr{K}$ è chiuso allora $p_{ij}^{(n)} = 0$, $\forall i \in \mathscr{K}$, $\forall j \notin \mathscr{K}$. Infatti, per induzione, se la proprietà è vera per n allora

$$p_{ij}^{(n+1)} = \sum_{k=1}^{N} p_{ik} p_{kj}^{(n)} = \sum_{k \in \mathscr{K}} p_{ik} p_{kj}^{(n)} = 0$$

in quanto $p_{ik} = 0$ se $k \notin \mathscr{K}$ e $p_{ij}^{(n)} = 0$ per $k \in \mathscr{K}$ per ipotesi. Sia $i \in \mathscr{K} \cap \mathscr{C}_\nu$; se $j \in \mathscr{C}_\nu \setminus \mathscr{K}$ allora $p_{ij}^{(n)} = 0$ per ogni n, ma allora i e j non sarebbero intercomunicanti; quindi $\mathscr{C}_\nu \subset \mathscr{K}$. ∎

Dalla proposizione 1 consegue che ogni insieme chiuso $\mathscr{K}$ è l'unione di classi di equivalenza

$$\mathscr{K} = \mathscr{C}_{\nu_1} \cup \mathscr{C}_{\nu_2} \cup \ldots \cup \mathscr{C}_{\nu_k}$$

Lemma 2 *Se* $\mathscr{S} = \mathscr{C}_1 \cup \mathscr{C}_2 \cup \cdots \cup \mathscr{C}_M$ *con* $M > 1$, *allora esiste un insieme chiuso* $\mathscr{K} \neq \mathscr{S}$.

Dimostrazione: È chiaro, a meno di una diversa numerazione, che per ogni $i \in \mathscr{C}_1$ e ogni $j \in \mathscr{C}_2$ risulta che $p_{ij}^{(n)} = 0$ per ogni n: infatti ci si convince facilmente che esiste almeno un $\underline{i} \in \mathscr{C}_1$ e un $\underline{j} \in \mathscr{C}_2$ tali che $p_{\underline{i}\,\underline{j}}^{(n)} = 0$ per ogni n; se $i \in \mathscr{C}_1$ e $j \in \mathscr{C}_2$ esistono $\bar{n}$ e $\bar{m}$ tali che

$$p_{\underline{i}i}^{(\bar{n})} > 0, \qquad p_{j\underline{j}}^{(\bar{m})} > 0$$

e quindi $p_{\underline{i}\,\underline{j}}^{(\bar{n}+\bar{m}+n)} \geqslant p_{\underline{i}i}^{(\bar{n})} p_{ij}^{(n)} p_{j\underline{j}}^{(\bar{m})}$ implica che $p_{ij}^{(n)} = 0$ per ogni n.

Consideriamo allora

$$\mathscr{K} = \mathscr{C}_1 \cup \mathscr{C}_{\nu_1} \cup \mathscr{C}_{\nu_2} \cup \ldots \cup \mathscr{C}_{\nu_p}$$

dove $\mathscr{C}_{\nu_s}$ sono <u>tutte</u> quelle classi che godono della stessa proprietà di $\mathscr{C}_1$: cioè $\forall i \in \mathscr{C}_{\nu_s}$, $\forall j \in \mathscr{C}_2 \Rightarrow p_{ij}^{(n)} = 0$ per ogni n. Sia $k \in \mathscr{K}$ e $i \notin \mathscr{K}$; i gode della proprietà che esiste un $\bar{n}$ tale che $p_{ji}^{(\bar{n})} > 0$ per almeno un $j \in \mathscr{C}_2$; quindi

$$p_{kj}^{(\bar{n}+1)} \geqslant p_{ki} p_{ij}^{(\bar{n})}$$

implica che $p_{ki} = 0$ (dato che $p_{kj}^{(\bar{n}+1)} = 0$). ∎

Proposizione 2 *Le due proprietà seguenti sono equivalenti:*

α) $\mathscr{S}$ e $\emptyset$ sono gli unici insiemi chiusi.

β) Esiste una sola classe di equivalenza ($M = 1$).

Dimostrazione: Il lemma 2 e la proposizione 2 forniscono facilmente la dimostrazione. ∎

Definizione 4 *Una catena di Markov si dice* irriducibile *se una delle proprietà α) o β) è verificata.*

4.1.1 Periodicità

Diremo che una stato i ha periodo $d \geq 1$ se accade

1) $p_{ii}^{(n)} = 0$ *se n* non *è un multiplo di d,*

2) *d è il più grande tra i numeri δ che godono della proprietà 1)* (con δ al posto di d)

Proposizione 3 *Risulta che*

a) i ha periodo d se e solo se $d = \text{M.C.D.}\{n \geq 1 : p_{ii}^{(n)} > 0\}$,

b) se i e j sono in una medesima classe di equivalenza, allora hanno lo stesso periodo.

Dimostrazione: La 1) implica che se $p_{ii}^{(n)} > 0$ allora d divide n e quindi divide il M.C.D.$\{n \geq 1 : p_{ii}^{(n)} > 0\}$; d'altra parte $\delta =$M.C.D.$\{n \geq 1 : p_{ii}^{(n)} > 0\}$ soddisfa la proprietà 1): quest'ultima affermazione implica che $\delta \leq d$, mentre la prima affermazione implica che $d \leq \delta$, da cui la a).
Infine se i e j sono in una medesima classe di equivalenza esistono due interi n, m tali che

$$p_{ij}^{(n)} > 0 \qquad \text{e} \qquad p_{ji}^{(m)} > 0$$

Se i ha periodo d allora per qualche s (multiplo di d) si ha $p_{ii}^{(s)} > 0$ e $p_{ii}^{(2s)} \geqslant p_{ii}^{(s)} p_{ii}^{(s)} > 0$; dalle disuguaglianze

$$p_{jj}^{(n+m+s)} \geqslant p_{ji}^{(m)} p_{ii}^{(s)} p_{ij}^{(n)} > 0 \qquad p_{jj}^{(n+m+2s)} \geqslant p_{ji}^{(m)} p_{ii}^{(2s)} p_{ij}^{(n)} > 0$$

ricaviamo che il periodo d_1 di j divide sia $n+m+s$ che $n+m+2s$; quindi d_1 divide la loro differenza s. Osserviamo allora che se d_1 divide s allora divide anche d; d'altra parte scambiando i ruoli di i e j abbiamo che $d \equiv d_1$. ∎

Possiamo associare allora un periodo d ad ogni classe di equivalenza, prendendo il periodo di un qualsiasi suo elemento.

Consideriamo una classe di equivalenza $\mathscr{C} \subset \mathscr{S}$ con periodo d. Siano $i, j \in \mathscr{C}$

Lemma 3 *Tutti gli interi n tali che* $p_{ij}^{(n)} > 0$ *appartengono ad una stessa classe resto* modulo *d.*

Dimostrazione: Infatti se n, n' sono tali che $p_{ij}^{(n)} > 0$, $p_{ij}^{(n')} > 0$, ricordando la seconda delle (7), abbiamo

$$p_{ii}^{(m+n)} \geqslant p_{ij}^{(n)} p_{ji}^{(m)} > 0 \qquad p_{ii}^{(m+n')} \geqslant p_{ij}^{(n')} p_{ji}^{(m)} > 0;$$

quindi, dovendo d dividere sia $m + n$ che $m + n'$, d divide anche la loro differenza $n - n'$. ∎

Tenendo fisso i possiamo ripartire gli elementi $j \in \mathscr{C}$ in classi $\mathscr{C}_\nu(i)$, $0 \leq \nu \leq d - 1$, dove ogni $\mathscr{C}_\nu(i)$ è formata da quegli elementi j, tali che, se $p_{ij}^{(n)} > 0$, allora $n \equiv \nu \pmod{d}$.

Sia $0 \leq r \leq d-1$; consideriamo $\mathscr{C}_r(j)$. Se $k \in \mathscr{C}_r(j)$ allora $p_{jk}^{(m)} > 0$ per qualche $m \equiv r(\mathsf{mod}\ d)$; allora

$$p_{ik}^{(n+m)} \geqslant p_{ij}^{(n)} p_{jk}^{(m)} > 0 \qquad \text{se } j \in \mathscr{C}_\nu(i)$$

con $n \equiv \nu(\mathsf{mod}\ d)$; quindi $n+m \equiv r+\nu(\mathsf{mod}\ d)$ da cui

$$\mathscr{C}_{r+\nu}(i) = \mathscr{C}_r(j) \qquad \text{se } j \in \mathscr{C}_\nu(i). \tag{8}$$

Ne consegue che $\{\mathscr{C}_r(j), 0 \leq r \leq d-1\} \equiv \{\mathscr{C}_r(i), 0 \leq r \leq d-1\}$ (come insiemi, dato che la (9) esprime che le classi si scambiano come permutazioni cicliche).

4.1.2 Ricorrenza e non-ricorrenza

Definizione 5 *Sia*

$$f_{ij}^{(n)} = \mathbb{P}(X_n = j, X_{n-1} \neq j, X_{n-2} \neq j, \ldots, X_1 \neq j | X_0 = i) \tag{9}$$

la probabilità di passare per j $\boxed{\textit{per la prima volta}}$ *dopo n unità temporali partendo da i. In particolare*

$$f_{ii}^{(n)} = \mathbb{P}(X_n = i, X_{n-1} \neq i, X_{n-2} \neq i, \ldots, X_1 \neq i | X_0 = i) \tag{10}$$

è la probabilità di tornare in i per la prima volta dopo n passi (unità temporali) ([4]).

Proposizione 4 *Risulta*

$$\sum_{n=1}^{\infty} f_{ii}^{(n)} \leq 1, \qquad \sum_{n=1}^{\infty} f_{ij}^{(n)} \leq 1 \tag{11}$$

inoltre vale la seguente relazione

$$p_{ij}^{(n)} = \sum_{\nu=1}^{n} f_{ij}^{(\nu)} p_{jj}^{(n-\nu)} \qquad \left(p_{ii}^{(n)} = \sum_{\nu=1}^{n} f_{ii}^{(\nu)} p_{ii}^{(n-\nu)}\right) \tag{12}$$

con $n \geq 1$.

[4]All'occasione porremo $f_{ij}^{(0} = 0$, in particolare $f_{ii}^{(0} = 0$.

Dimostrazione: Per la (11) basta osservare che gli insiemi

$$(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j), \quad (X_{n+k} = j, X_{n+k-1} \neq j, \ldots, X_1 \neq j)$$

sono disgiunti. Per quanto riguarda la (12), consideriamo

$$(X_n = j) = \bigcup_{\nu=1}^{n} (X_n = j, X_\nu = j, X_{\nu-1} \neq j, \ldots, X_1 \neq j) \tag{13}$$

Osserviamo che l'unione (13) è disgiunta e che

$$\begin{aligned}
&\mathbb{P}(X_n = j, X_\nu = j, X_{\nu-1} \neq j, \ldots, X_1 \neq j | X_0 = i) = \\
&\qquad = \mathbb{P}(X_n = j | X_\nu = j, X_{\nu-1} \neq j, \ldots, X_1 \neq j, X_0 = i) \cdot \\
&\qquad \cdot \mathbb{P}(X_\nu = j, X_{\nu-1} \neq j, \ldots, X_1 \neq j | X_0 = i) = \\
&= \mathbb{P}(X_n = j | X_\nu = j) \cdot \mathbb{P}(X_\nu = j, X_{\nu-1} \neq j, \ldots, X_1 \neq j | X_0 = i) = p_{jj}^{(n-\nu)} \, f_{ij}^{(\nu)}
\end{aligned}$$

La (12) si ricava combinando quest'ultima relazione con la (13). ∎

Nota. Osserviamo che

$$\sum_{n=1}^{M} p_{ij}^{(n)} = \sum_{n=1}^{M} \sum_{\nu=1}^{n} f_{ij}^{(\nu)} p_{jj}^{(n-\nu)} = \sum_{\nu=1}^{M} f_{ij}^{(\nu)} \sum_{n=0}^{M-\nu} p_{jj}^{(n)} \quad \left(\leq \sum_{\nu=1}^{M} f_{ij}^{(\nu)} \cdot \sum_{n=0}^{M} p_{jj}^{(n)} \right) \tag{14}$$

Proposizione 5 *Le seguenti affermazioni sono equivalenti*

$$\alpha) \qquad \sum_{n=0}^{\infty} p_{ii}^{(n)} = +\infty \qquad \left(\sum_{n=0}^{\infty} p_{ii}^{(n)} < +\infty \right)$$

$$\beta) \qquad \sum_{n=1}^{\infty} f_{ii}^{(n)} = 1 \qquad \left(\sum_{n=1}^{\infty} f_{ii}^{(n)} < 1 \right)$$

Dimostrazione: Supponiamo $\sum_{n=0}^{\infty} p_{ii}^{(n)} < +\infty$; dalla (14), per M tendente all'infinito, otteniamo (prendendo $j = i$)

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \sum_{\nu=1}^{\infty} f_{ii}^{(\nu)} \cdot \sum_{n=0}^{\infty} p_{ii}^{(n)}$$

da cui

$$\sum_{\nu=1}^{\infty} f_{ii}^{(\nu)} = 1 - \frac{1}{\sum_{n=0}^{\infty} p_{ii}^{(n)}} < 1$$

Viceversa, se $\sum_{\nu=1}^{\infty} f_{ii}^{(\nu)} < 1$, dalla (14) otteniamo

$$\sum_{n=0}^{M} p_{ii}^{(n)} \leq \frac{1}{1-\sum_{\nu=1}^{M} f_{ii}^{(\nu)}} \leq \frac{1}{1-\sum_{\nu=1}^{\infty} f_{ii}^{(\nu)}},$$

che implica la convergenza della serie $\sum_{n=0}^{\infty} p_{ii}^{(n)}$. ∎

Introduciamo la seguente classificazione:

Definizione 6 *Denotiamo con* $f_{ii}^* = \sum_{n=1}^{\infty} f_{ii}^{(n)}$ (risp. $f_{ij}^* = \sum_{n=1}^{\infty} f_{ij}^{(n)}$) *la probabilità di ritornare in* i *(nel futuro)* (risp. *la probabilità di raggiungere* j *nel futuro).*
Diremo che lo stato i *è* ricorrente *se* $f_{ii}^* = 1$*; diremo che lo stato* i *è* transiente *(non ricorrente) se* $f_{ii}^* < 1$.

Proposizione 6 *Se* i, j *appartengono alla stessa classe di equivalenza* $\mathscr{C}$ *allora sono entrambi o ricorrenti oppure entrambi transienti.*

Dimostrazione: Se i, j appartengono alla stessa classe di equivalenza $\mathscr{C}$ valgono le (7). Consideriamo

$$p_{jj}^{(n+m+s)} \geqslant p_{ji}^{(m)} p_{ii}^{(s)} p_{ij}^{(n)}$$

da cui

$$\sum_{k=0}^{\infty} p_{jj}^{(k)} \geqslant \sum_{s=0}^{\infty} p_{jj}^{(n+m+s)} \geqslant p_{ji}^{(m)} \Big(\sum_{s=0}^{\infty} p_{ii}^{(s)} \Big) p_{ij}^{(n)}.$$

Quest'ultima relazione mostra che se i è ricorrente anche j è ricorrente; scambiando il ruolo di j con il ruolo di i otteniamo la dimostrazione. ∎

Con quest'ultima proposizione potremo definire che una classe di equivalenza è ricorrente (risp. transiente) quando un qualsiasi suo elemento (e quindi tutti) è ricorrente (risp. transiente).
Scriveremo $\mathscr{S} = \mathscr{T}_1 \cup \mathscr{T}_2 \cup \cdots \cup \mathscr{T}_t \cup \mathscr{R}_1 \cup \mathscr{R}_2 \cup \cdots \cup \mathscr{R}_r$ *dove* $\mathscr{T}_i$, $\mathscr{R}_i$ *sono classi di equivalenza rispettivamente transienti e ricorrenti.*

Proposizione 7 *Una classe di equivalenza ricorrente è chiusa.*

Dimostrazione: Sia $\mathscr{C}$ la classe di equivalenza ricorrente. Sia, per assurdo, $i \in \mathscr{C}$ e $j \notin \mathscr{C}$ tale che $p_{ij} > 0$; osserviamo che $p_{jk}^{(m)} = 0$ per ogni m e per ogni $k \in \mathscr{C}$ (altrimenti, se n è un intero tale che $p_{ki}^{(n)} > 0$, avremo $p_{ji}^{(n+m)} \geqslant p_{jk}^{(m)} p_{ki}^{(n)} > 0$).

Consideriamo $\mathbb{P}(X_n \notin \mathscr{C}\,|X_1 = j)$; tenendo presente che $p_{jk} = 0$ per ogni $k \in \mathscr{C}$ e l'equazione di Chapman-Kolmogorov, abbiamo:

$$\sum_{k\notin\mathscr{C}} p_{jk}^{(n-1)} = \mathbb{P}(X_n \notin \mathscr{C}\,|X_1 = j) = \sum_{k_2\notin\mathscr{C}} \sum_{k_3\notin\mathscr{C}} \cdots \sum_{k_n\notin\mathscr{C}} p_{jk_2}p_{k_2k_3}\cdots p_{k_{n-1}k_n} = 1$$

dato che $\mathbb{P}(X_n \in \mathscr{C}\,|X_1 = j) = \sum_{k\in\mathscr{C}} p_{jk}^{(n-1)} = 0$. Consideriamo ora

$$\begin{aligned}
p_{ij} = p_{ij}\mathbb{P}(X_n \notin \mathscr{C}\,|X_1 = j) &\leqslant \sum_{k_1\notin\mathscr{C}} p_{ik_1} \sum_{k_2\notin\mathscr{C}} \cdots \sum_{k_n\notin\mathscr{C}} p_{k_1k_2}\cdots p_{k_{n-1}k_n} \\
&\leqslant \sum_{k_1\neq i}\sum_{k_2\neq i}\cdots\sum_{k_n\neq i} p_{ik_1}p_{k_1k_2}\cdots p_{k_{n-1}k_n} \\
&= \mathbb{P}\Big(\bigcap_{\nu=1}^{n}(X_\nu \neq i)|X_0 = i\Big)
\end{aligned}$$

Denotiamo con $B_n = \bigcap_{\nu=1}^n (X_\nu \neq i)$; abbiamo $B_n \downarrow B = \bigcap_{\nu=1}^\infty (X_\nu \neq i)$. D'altra parte è facile verificare, denotando con

$$A = \bigcup_{n=1}^{\infty} (X_n = i, X_{n-1} \neq i, \ldots, X_1 \neq i)$$

che $A^c = B$. Ma se $\mathscr{C}$ è ricorrente allora $\mathbb{P}(A|X_0 = i) = \sum_{n=1}^\infty f_{ii}^{(n)} = f_{ii}^* = 1$; quindi $\mathbb{P}(B|X_0 = i) = 0$. Riassumendo abbiamo

$$0 < p_{ij} \leqslant \mathbb{P}(B_n|X_0 = i) \downarrow \mathbb{P}(B|X_0 = i) = 0$$

che è manifestamente un assurdo: quindi $p_{ij} = 0$. ∎

Proposizione 8 *Una classe di equivalenza transiente con un numero finito di elementi non è chiusa.*

Dimostrazione: Sia $\mathscr{C}$ la classe di equivalenza transiente. Se $\mathscr{C}$ fosse chiusa, per ogni $i \in \mathscr{C}$ e ogni $j \notin \mathscr{C}$ si avrebbe che $p_{ij}^{(n)} = 0$ per ogni n. D'altra parte, visto che $(p_{ij}^{(n)})$ è una matrice stocastica, abbiamo per ogni n

$$\sum_{k\in\mathscr{C}} p_{ik}^{(n)} = 1. \tag{15}$$

In secondo luogo, se k è uno stato transiente, dalla (14) ricaviamo (per $M \to \infty$)

$$\sum_{n=1}^{\infty} p_{ik}^{(n)} \leqslant \sum_{\nu=1}^{\infty} f_{ik}^{(\nu)} \sum_{n=0}^{\infty} p_{kk}^{(n)} < \infty$$

da cui $p_{ik}^{(n)} \to 0$ per ogni $i \in \mathscr{C}$. Quindi passando al limite nella (15) otteniamo l'assurdo $0 = 1$! ∎

Corollario 1 *Se j è uno stato transiente allora $p_{ij}^{(n)} \to 0$ per $n \to +\infty$ e per ogni $i \in \mathscr{S}$.*

4.2 Tempo medio di passaggio (primo passaggio)

Se i è uno stato ricorrente sappiamo che risulta $f_{ii}^* = \sum_{n=1}^{\infty} f_{ii}^{(n)} = 1$. Possiamo allora *pensare* $\{f_{ii}^{(n)}\}$ come la distribuzione di probabilità di una variabile casuale T_i a valori in $\mathbb{N}$, cioè

$$\mathbb{P}(T_i = n) = f_{ii}^{(n)}$$

T_i si *interpreta* come il tempo in cui accade il primo passaggio (oppure il primo ritorno) in i, partendo da i. Consideriamo allora la media

$$\mathbb{E}(T_i) = \sum_{n=1}^{\infty} n f_{ii}^{(n)} =: \mu_i \tag{16}$$

Definizione 7 *Se risulta $\mu_i = \mathbb{E}(T_i) < +\infty$ diremo che lo stato ricorrente i è di tipo* positivo*; nel caso in cui $\mu_i = +\infty$ diremo che lo stato ricorrente i è di tipo* nullo.

Riassumiamo infine dicendo che lo spazio degli stati $\mathscr{S}$ di una catena di Markov si divide in $M = t + r$ classi di equivalenza di cui t transienti e r ricorrenti

$$\mathscr{S} = \mathscr{T}_1 \cup \mathscr{T}_2 \cup \cdots \cup \mathscr{T}_t \cup \mathscr{R}_1 \cup \mathscr{R}_2 \cup \cdots \cup \mathscr{R}_r$$

Ad ogni classe è associato un periodo: in particolare sia d_ν il periodo della classe ricorrente $\mathscr{R}_\nu$, $1 \le \nu \le r$. Ciascuna classe ricorrente $\mathscr{R}_\nu$ si partiziona in d_ν sottoinsiemi $\mathscr{C}_{\nu\mu}$ con $0 \le \mu \le d_\nu - 1$. Infine ciascuna classe $\mathscr{R}_\nu$ è chiusa e ad ogni $i \in \mathscr{R}_\nu$ è associato μ_i.

Osservazione. La catena di Markov si può restringere ad ogni insieme chiuso di $\mathscr{S}$ rimanendo ancora una catena di Markov, i cui stati sono in quel sottoinsieme chiuso. In particolare, ciascuna catena di Markov, ottenuta restringendo l'originaria catena di Markov a $\mathscr{R}_\nu$, è una catena irriducibile.

Enunciamo ora il teorema sul comportamento asintotico delle $p_{ij}^{(n)}$ al variare degli stati i e j. Alcune affermazioni sono state dimostrate precedentemente.

Teorema 1 *Data una catena di Markov con spazio degli stati $\mathscr{S} = \{1, 2, \ldots, N\}$ risulta:*

1. *Se j è uno stato transiente allora $p_{ij}^{(n)} \to 0$ per ogni $i \in \mathscr{S}$.*

2. *Se j è uno stato ricorrente con periodo d_j* ($\equiv d_\nu$ della classe $\mathscr{R}_\nu$ a cui j appartiene ([5])) *e tempo medio di primo passaggio μ_i, allora*

   a) se $i \in \mathscr{R}_\ell \neq \mathscr{R}_\nu$, allora $p_{ij}^{(n)} = 0$ per ogni n,

   b) se $i \in \mathscr{R}_\nu$, allora, se $j \in \mathscr{C}_{\nu\mu}$, abbiamo

   $$\lim_{n\to+\infty} p_{ij}^{(nd_j+\mu)} = \frac{d_j}{\mu_j}, \qquad (= 0 \textit{ se } \mu_j = +\infty) \qquad (b')$$

   e

   $$p_{ij}^{(n)} = 0 \quad \textit{se } n \not\equiv \mu \ (\mathsf{mod}\ d_j)$$

   c) se i è transiente, allora per ogni $0 \leq \bar\mu \leq d_j - 1$

   $$\lim_{n\to+\infty} p_{ij}^{(nd_j+\bar\mu)} = f_{ij}^*(\bar\mu)\,\frac{d_j}{\mu_j}$$

   dove

   $$f_{ij}^*(\bar\mu) = \sum_{\substack{n=1 \\ n\equiv\mu\ (\mathsf{mod}\ d_j)}}^{\infty} f_{ij}^{(n)}. \ ([6])$$

Corollario 2 *Se $(X_n, n \geq 0)$ è una catena di Markov irriducibile e aperiodica ($d = 1$) risulta che*

$$p_{ij}^{(n)} \to \pi_j = \frac{1}{\mu_j}$$

In tal caso

$$P^n \to Q = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_N \\ \pi_1 & \pi_2 & \cdots & \pi_N \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_N \end{pmatrix}$$

Osservare che

$$\pi_1 + \pi_2 + \cdots + \pi_N = 1 = \frac{1}{\mu_1} + \frac{1}{\mu_2} + \cdots + \frac{1}{\mu_N}.$$

Inoltre, da $P^nP = PP^n = P^{n+1}$, passando al limite otteniamo $QP = PQ = Q$; in altre parole

$$(\pi_1, \pi_2, \cdots, \pi_N)P = (\pi_1, \pi_2, \cdots, \pi_N)$$

[5] Sia μ tale che $j \in \mathscr{C}_{\nu\mu}$, con $0 \leq \mu \leq d_\nu - 1$.

[6] Osservare che $\sum_{\bar\mu=0}^{d_j-1} f_{ij}^*(\bar\mu) = f_{ij}^*$

cioè $(\pi_1, \pi_2, \cdots, \pi_N)$ è l'autovettore (sinistro e normalizzato) associato all'autovalore 1 di P (autovettore destro di P^*). Questa osservazione porge un metodo elementare per ricavarsi $(\pi_1, \pi_2, \cdots, \pi_N)$, che è chiamata la distribuzione stazionaria per una catena di Markov irriducibile e aperiodica.
Nel caso in cui lo spazio degli stati è numerabile oltre a essere irriducibile e aperiodica bisogna richiedere che la catena di Markov sia ricorrente.

Consideriamo la famiglia $\mathscr{P}$ delle cosiddette matrici stocastiche (anche infinite), cioè le matrici P tali che

$$p_{ij} \geq 0 \quad \forall i,j \qquad \textit{and} \qquad \sum_{j=1}^{N} p_{ij} = 1 \quad \forall i$$

dove la somma diventa una serie nel caso di matrici infinite.
Esercizio. $I \in \mathscr{P}$; se $P, Q \in \mathscr{P}$ allora $PQ \in \mathscr{P}$.
Quindi se P è una matrice stocastica allora lo sono tutte le sue potenze!

$$I, P, P^2, P^3, P^4, \ldots, P^n, \ldots$$

e indicheremo gli elementi di matrice corrispondenti con

$$p_{ij}^{(0)} = \delta_{ij}, p_{ij}^{(1)} = p_{ij}, p_{ij}^{(2)}, p_{ij}^{(3)}, p_{ij}^{(4)}, \ldots, p_{ij}^{(n)}, \ldots$$

per esempio $p_{ij}^{(2)} = \sum_k p_{ik}p_{kj}$.

Sia $\mathscr{S} = \{1, 2, 3, \ldots, N\}$ nel caso di matrici finite, oppure $\mathscr{S} = \mathbb{N}$ oppure $\mathscr{S} = \mathbb{Z}$ nel caso di matrici infinite.
Definizione. *Data la matrice stocastica P, diremo che $i \sim j$, dove $i, j \in \mathscr{S}$, se esiste una coppia n', n'' di interi non negativi tali che*

$$p_{ij}^{(n')} > 0 \qquad \textit{and} \qquad p_{ij}^{(n'')} > 0$$

Esercizio. La relazione $\sim$ è una relazione di equivalenza in $\mathscr{S}$. Quindi dà luogo ad una partizione di $\mathscr{S}$.
Definizione. *Diremo che la matrice stocastica P è irriducibile se la partizione associata è formata da un unico insieme (cioè l'insieme $\mathscr{S}$).*
Definizione. *Data la matrice stocastica P, definiamo periodo d_i di ogni $i \in \mathscr{S}$ il massimo comun divisore di tutti gli $n \geq 1$ tali che $p_{ii}^{(n)} > 0$.*

Osservazione. In maniera esplicita si ha

$$p_{ii}^{(n)} = \sum_{k_1, k_2, \ldots, k_{n-1}} p_{ik_1}p_{k_1k_2} \cdots p_{k_{n-2}k_{n-1}}p_{k_{n-1}i}$$

Definiamo

$$f_{ii}^{(n)} = \sum_{k_1 \neq i, k_2 \neq i, \ldots, k_{n-1} \neq i} p_{ik_1} p_{k_1 k_2} \cdots p_{k_{n-2} k_{n-1}} p_{k_{n-1} i}$$

Sussistono le seguenti identità

$$\sum_{k=1}^{n} f_{ii}^{(k)} = 1 - \sum_{k_1 \neq i, k_2 \neq i, \ldots, k_n \neq i} p_{ik_1} p_{k_1 k_2} \cdots p_{k_{n-1} k_n}$$

$$p_{ii}^{(n)} = \sum_{k=1}^{n} f_{ii}^{(k)} p_{ii}^{(n-k)}$$

Dalla prima identità segue che

$$f_{ii}^{*} \equiv \sum_{k=1}^{\infty} f_{ii}^{(k)} \leq 1.$$

Dalla seconda segue che, definendo $q_n = \sum_{k=0}^{n} p_{ii}^{(k)}$ e scambiando le somme,

$$q_n = 1 + \sum_{k=1}^{n} f_{ii}^{(k)} q_{n-k}$$

da cui

$$q_n \leq 1 + q_n \sum_{k=1}^{n} f_{ii}^{(k)}$$

cioè

$$q_n (1 - \sum_{k=1}^{n} f_{ii}^{(k)}) \leq 1$$

da cui segue che se $q_n \nearrow +\infty$ allora $\sum_{k=1}^{\infty} f_{ii}^{(k)} = 1$. D'altra parte se $q_n \nearrow q$ allora risulta che, fissato m,

$$q = 1 + q \sum_{k=1}^{m} f_{ii}^{(k)} + \lim_{n \to \infty} \Big(\sum_{k=m+1}^{n} f_{ii}^{(k)} q_{n-k} \Big)$$

da cui

$$\sum_{k=1}^{m} f_{ii}^{(k)} = 1 - \frac{1 + \lim_{n \to \infty} \Big(\sum_{k=m+1}^{n} f_{ii}^{(k)} q_{n-k} \Big)}{q}$$

Ci si convince subito che

$$\lim_{m\to\infty}\lim_{n\to\infty}\Big(\sum_{k=m+1}^{n} f_{ii}^{(k)} q_{n-k}\Big) = 0$$

e quindi otteniamo

$$\sum_{k=1}^{\infty} f_{ii}^{(k)} = 1 - \frac{1}{q} < 1$$

Abbiamo quindi che

$$\sum_{k=1}^{\infty} f_{ii}^{(k)} = 1 \Longleftrightarrow \sum_{k=0}^{\infty} p_{ii}^{(k)} = \infty$$

Definizione *Diremo che* $i \in \mathscr{S}$ *è* ricorrente *se* $\sum_{k=1}^{\infty} f_{ii}^{(k)} = 1$. *Diremo che* $i \in \mathscr{S}$ *è* transiente *se* $\sum_{k=1}^{\infty} f_{ii}^{(k)} < 1$.

Consideriamo

$$1 - \sum_{k=1}^{n} f_{jj}^{(k)} = \sum_{\substack{k_1\neq j\\ k_2\neq j\\ \vdots\\ k_n\neq j}} p_{jk_1}p_{k_1k_2}\cdots p_{k_{n-1}k_n}$$

e sia $m \geq 1$ il primo intero tale che $p_{ji}^{(m)} > 0$; allora per $0 < \nu < m$

$$p_{ji}^{(m)} = \sum_{k} p_{jk}^{(\nu)} p_{ki}^{(m-\nu)} = \sum_{k\neq j} p_{jk}^{(\nu)} p_{ki}^{(m-\nu)} = \sum_{\substack{k_1\neq j\\ k_2\neq j\\ \vdots\\ k_{m-1}\neq j}} p_{jk_1}p_{k_1k_2}\cdots p_{k_{m-1}i}$$

allora

$$1 - \sum_{k=1}^{n} f_{jj}^{(k)} \geq p_{ji}^{(m)} \sum_{\substack{k_{m+1}\neq j\\ k_{m+2}\neq j\\ \vdots\\ k_n\neq j}} p_{jk_{m+1}}p_{k_{m+1}k_{m+2}}\cdots p_{k_{n-1}k_n} = p_{ji}^{(m)}(1 - \sum_{k=1}^{n-m} f_{ij}^{(k)})$$

da cui
$$1-f_{jj}^{*}\geq p_{ji}^{(m)}\,(1-f_{ij}^{*})$$

Lemma 4 *Sia d=M.C.D.$\{n>0: p_{ii}^{(n)}>0\}$ e sia d'=M.C.D.$\{n>0: f_{ii}^{(n)}>0\}$ allora $d=d'$.*

Dimostrazione: Risulta che

$$p_{ii}^{(n)}=\sum_{k=1}^{n} f_{ii}^{k)}p_{ii}^{(n-k)}\geq f_{ii}^{(n)} \tag{1}$$

da cui
$$\mathscr{F}=\{n>0: f_{ii}^{(n)}>0\}\subset\mathscr{P}=\{n>0: p_{ii}^{(n)}>0\}$$

Sia $\tilde{\mathscr{F}}=\{n=kd', k=1,2,3\cdots\}$; risulta $\mathscr{F}\subset\tilde{\mathscr{F}}$ ed entrambi hanno lo stesso massimo comun divisore. Supponiamo, per assurdo, che $\mathscr{P}$ non sia un sottoinsieme di $\tilde{\mathscr{F}}$: in tal caso esiste un primo n tale che $p_{ii}^{(n)}>0$ ed n non è congruo a zero modulo d'. D'altra parte dalla (1) ricaviamo che deve esistere un $\nu<n$ tale che
$$f_{ii}^{(\nu)}p_{ii}^{(n-\nu)}>0$$
da cui $n\equiv 0 \bmod d'$.Hence

$$\mathscr{F}\subset\mathscr{P}\subset\tilde{\mathscr{F}}$$

ed il lemma è dimostrato. ∎

Lemma 5
$$f_{ij}^{*}=\sum_{k=1}^{\infty} f_{ij}^{(k)}=1$$

Dimostrazione: Dati i,j due stati di una classe ricorrente, esiste un intero n minimo tale che $p_{ji}^{(n)}>0$ (e quindi $p_{ji}^{(k)}=0$ per $k\leq n-1$). Ebbene

$$p_{ji}^{(n)}=\mathbb{P}(X_n=i,\bigcap_{\ell=1}^{n-1}(X_\ell\neq j)|X_0=j)+\mathbb{P}(X_n=i,\bigcup_{\ell=1}^{n-1}(X_\ell=j)|X_0=j).$$

Ora si ha

$$\mathbb{P}(X_n=i,\bigcup_{\ell=1}^{n-1}(X_\ell=j)|X_0=j)\leqslant\sum_{\ell=1}^{n-1}\mathbb{P}(X_n=i,X_\ell=j|X_0=j)$$

ma

$$\mathbb{P}(X_n = i, X_\ell = j | X_0 = j) = P(X_\ell = j | X_0 = j) \cdot \mathbb{P}(X_n = i | X_\ell = j)$$

da cui, ricordando che $\mathbb{P}(X_n = i | X_\ell = j) = 0$

$$p_{ji}^{(n)} = \mathbb{P}(X_n = i, \bigcap_{\ell=1}^{n-1} (X_\ell \neq j) | X_0 = j) > 0. \tag{$*$}$$

Ancora

$$0 = 1 - f_{jj}^* = \mathbb{P}(\bigcap_{\ell=1}^{\infty} (X_\ell \neq j) | X_0 = j) =$$

$$\sum_{k \neq j} \mathbb{P}(\bigcap_{\ell=n+1}^{\infty} (X_\ell \neq j) | X_n = k) \cdot \mathbb{P}(X_n = k, \bigcap_{\ell=1}^{n-1} (X_\ell \neq j) | X_0 = j) \geq$$

$$\mathbb{P}(\bigcap_{\ell=n+1}^{\infty} (X_\ell \neq j) | X_n = i) \cdot \mathbb{P}(X_n = i, \bigcap_{\ell=1}^{n-1} (X_\ell \neq j) | X_0 = j)$$

$$(1 - f_{ij}^*) \cdot \mathbb{P}(X_n = i, \bigcap_{\ell=1}^{n-1} (X_\ell \neq j) | X_0 = j)$$

da cui, tenendo presente $(*)$, abbiamo il lemma. ∎

5 Teorema limite

Sia $\{f_1, f_2, f_3, \ldots, f_n, \ldots\}$ una successione di numeri reali non negativi di somma 1, in altre parole

$$f_i \geq 0, \quad \forall i = 1, 2, \ldots \qquad \sum_{i=1}^{\infty} f_i = 1;$$

definiamo in modo ricorsivo la successione $\{u_0, u_1, u_2, u_3, \ldots, u_n, \ldots\}$ nel modo seguente $u_0 = 1$ e

$$u_n = u_0 f_n + u_1 f_{n-1} + u_2 f_{n-2} + \cdots + u_{n-1} f_1; \tag{17}$$

e tale successione ha le seguenti proprietà (tramite un facile procedimento di induzione)

$$0 \leq u_n \leq 1$$

Teorema 2 *Se vale la*

$$M.C.D.\{n \geq 1, f_n > 0\} = 1 \tag{18}$$

allora

$$u_n \to \frac{1}{\mu}$$

dove

$$\mu = \sum_{n=1}^{\infty} n f_n$$

dove sottointendiamo $\frac{1}{\mu} = 0$ *se* $\mu = \infty$*.*

Per la dimostrazione del teorema introduciamo la seguente notazione:

$$\varrho_k = f_{k+1} + f_{k+1} + f_{k+1} + \cdots = \sum_{i=k+1}^{\infty} f_i$$

e quindi $\varrho_0 = 1$. Inoltre risulta

$$\mu = \sum_{n=1}^{\infty} n f_n = \sum_{k=0}^{\infty} \varrho_k.$$

e vale la seguente identità, per ogni $n \geq 0$,

$$\varrho_0 u_n + \varrho_1 u_{n-1} + \varrho_2 u_{n-2} + \cdots + \varrho_{n-1} u_1 + \varrho_n u_0 = 1 \tag{19}$$

Consideriamo

$$\eta = \limsup_{n\to\infty} u_n$$

Applicando il risultato dell'appendice A esiste una successione di indici ν_1, ν_2, ν_3, ... tale che, per ogni $k = 0, 1, 2, \ldots$, risulta convergente la successione

$$u_{\nu_n - k} \to w_k$$

ad un limite w_k. Ovviamente $w_0 = \eta$ e per ogni $k = 0, 1, 2, \ldots$

$$0 \leq w_k \leq \eta.$$

D'altra parte dall'identità (17) otteniamo

$$u_{\nu_r - k} = \sum_{i=1}^{\infty} f_i u_{\nu_r - k - i}$$

dove in realtà la sommatoria è una somma finita. Passando al limite per $r \to \infty$ otteniamo

$$w_k = \sum_{i=1}^{\infty} f_i w_{k+i}$$

ed applicando i risultati dell'appendice B otteniamo che

$$w_0 = \eta = w_1 = w_2 = \dots$$

Supponiamo $\mu = \infty$. Dalla (19) deduciamo, fissato M

$$\varrho_0 u_N + \varrho_1 u_{N-1} + \varrho_2 u_{N-2} + \dots + \varrho_M u_{N-M} \leq 1$$

da cui passando al limite per $N \to \infty$, pensando N variabile nella successione $\nu_1, \nu_2, \nu_3, \dots$, otteniamo

$$\eta(\varrho_0 + \varrho_1 + \varrho_2 + \dots + \varrho_M) \leq 1$$

da cui $\eta = 0$. In questo caso il teorema è dimostrato.
Nel caso in cui $\mu < \infty$, fissato comunque ε e scegliendo M tale che

$$\sum_{i=M+1}^{\infty} \varrho_i < \varepsilon$$

oltre la disuguaglianza precedente, che continua a valere, abbiamo

$$1 - \varepsilon \leq \eta(\varrho_0 + \varrho_1 + \varrho_2 + \dots + \varrho_M) \leq 1$$

che implica $\boxed{\eta\mu = 1}$. Mostriamo ora che esiste il limite considerando

$$\eta_0 = \liminf_{n \to \infty} u_n.$$

prendiamo una sottosuccessione convergente ad η_0, indicandola con u_{n_k}. Allora scrivendo la (19), sempre fissando M, come

$$1 = \varrho_0 u_N + (\varrho_1 u_{N-1} + \dots + \varrho_M u_{N-M}) + (\varrho_{M+1} u_{N-M-1} + \dots + \varrho_N u_0)$$

abbiamo la disuguaglianza seguente

$$1 \leq \varrho_0 u_{n_k} + (\varrho_1 + \dots + \varrho_M)(\eta + \varepsilon) + (\varrho_{M+1} + \varrho_{M+2} + \cdots)$$

da cui, prima per $k \to \infty$ e successivamente per $M \to \infty$, otteniamo

$$1 \leq \varrho_0 \eta_0 + (\mu - \varrho_0)(\eta + \varepsilon);$$

ancora passando al limite per $\varepsilon \to 0$ e tenendo conto che $\mu\eta = 1$ otteniamo $\eta \leq \eta_0$. Quindi il $\liminf$ coincide con il $\limsup$ ed il teorema è dimostrato.

Teorema 3 *Sia $P = (p_{ij})$ la matrice (anche infinita) di una catena di Markov* irriducibile, ricorrente *e* aperiodica. *Allora $P^n \to Q$.*

Dimostrazione: Per quanto riguarda gli elementi diagonali $p_{ii}^{(n)}$ possiamo applicare il teorema precedente. Quindi

$$p_{ii}^{(n)} \to \frac{1}{\mu_i}.$$

Per gli elementi fuori dalla diagonale abbiamo

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

Da

$$\sum_{\nu=m+1}^{n} f_{ij}^{(\nu)} p_{jj}^{(n-\nu)} \leq \sum_{\nu=m+1}^{\infty} f_{ij}^{(\nu)},$$

fissato $\varepsilon > 0$, possiamo scegliere m in modo che

$$p_{ij}^{(n)} \leq \sum_{\nu=1}^{m} f_{ij}^{(\nu)} p_{jj}^{(n-\nu)} + \varepsilon,$$

da cui

$$\limsup_{n\to\infty} p_{ij}^{(n)} \leq \frac{1}{\mu_j} \sum_{\nu=1}^{m} f_{ij}^{(\nu)} + \varepsilon.$$

D'altra parte, per $n > m$, da

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

otteniamo

$$\frac{1}{\mu_j} \sum_{\nu=1}^{m} f_{ij}^{(\nu)} \leq \liminf_{n\to\infty} p_{ij}^{(n)} \leq \limsup_{n\to\infty} p_{ij}^{(n)} \leq \frac{1}{\mu_j} \sum_{\nu=1}^{m} f_{ij}^{(\nu)} + \varepsilon$$

che per $m \to \infty$ ed $\varepsilon \to 0$ diventa

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_j} \sum_{\nu=1}^{\infty} f_{ij}^{(\nu)}$$

e, per il lemma 4, il terorema è dimostrato. ∎

Appendice

A Un'applicazione del procedimento diagonale di Cantor.

Data una successione limitata di numeri reali $\{u_1, u_2, u_3, \ldots, u_n, \ldots\}$ possiamo estrarre una sottosuccessione convergente

$$u_{r_1}, u_{r_2}, u_{r_3}, \ldots u_{r_n}, \ldots \to w_0$$

Consideriamo la sottosuccessione (se $r_i - 1 \leq 0$ poniamo $u_{r_i - 1} = 0$)

$$u_{r_1-1}, u_{r_2-1}, u_{r_3-1}, \ldots u_{r_n-1}, \ldots$$

ed estraiamone una sottosuccessione convergente

$$u_{r_1^{(1)}-1}, u_{r_2^{(1)}-1}, u_{r_3^{(1)}-1}, \ldots u_{r_n^{(1)}-1}, \ldots \to w_1$$

Analogamente dalla successione

$$u_{r_1^{(1)}-2}, u_{r_2^{(1)}-2}, u_{r_3^{(1)}-2}, \ldots u_{r_n^{(1)}-2}, \ldots$$

estraiamone una sottosuccessione convergente

$$u_{r_1^{(2)}-2}, u_{r_2^{(2)}-2}, u_{r_3^{(2)}-2}, \ldots u_{r_n^{(2)}-2}, \ldots \to w_2$$

e via dicendo per ogni intero k. Allora la sottosuccessione

$$u_{\nu_1}, u_{\nu_2}, u_{\nu_3}, \ldots, u_{\nu_n}, \ldots$$

con $\nu_1 = r_1$ e $\nu_k = r_k^{(k-1)}$, per $k > 1$, è tale che

$$u_{\nu_k - k} \to w_k \qquad k = 0, 1, 2 \ldots$$

B Affermazione 2

Denotiamo con $\mathscr{N} = \{n \geq 1 : f_n > 0\}$ e con $\mathscr{N}^+$ tutti gli interi del tipo $p_1 n_1 + p_2 n_2 + \cdots + p_k n_k$ con $k \geq 1$ e per ogni $i = 1, 2 \ldots, k$

$$n_i \in \mathscr{N} \quad \text{e} \quad p_i \text{ è un intero positivo}$$

Ovviamente $\mathscr{N} \subset \mathscr{N}^+$.

Lemma 6 *Se vale la* (18)*, allora esiste un intero a partire dal quale tutti gli interi maggiori sono contenuti in* $\mathcal{N}^+$*.*

Proof: Per una proprietà nota del M.C.D. esistono $n_1, n_2, \ldots, n_k$ in $\mathcal{N}$ e $\nu_1, \nu_2, \ldots, \nu_k$ in $\mathbb{Z}$ tali che

$$n_1\nu_1 + n_2\nu_2 + \cdots + n_k\nu_k = 1.$$

Indichiamo con $s = n_1+n_2+\cdots+n_k$. Allora ogni intero n possiamo scriverlo come

$$n = qs + r = (q + \nu_1 r)n_1 + (q + \nu_2 r)n_2 + \cdots + (q + \nu_k r)n_k,$$

da cui, prendendo $q \geq \nu s$ dove $\nu = \max\{|\nu_i|,\ i = 1, 2, \ldots, k\} > 0$, ogni intero più grande di νs appartiene ad $\mathcal{N}^+$.

Lemma 7 *Se vale la* (18)*, data una successione* $\{w_0, w_1, w_2, \ldots, w_n, \ldots\}$ *tale che*

$$w_n = \sum_{k=1}^{\infty} f_k\, w_{n+k}, \tag{20}$$

per ogni $n \geq 0$*,* $0 \leq w_n \leq 1$ *e* $w_0 = 1$ *allora* $w_n = 1$ *per ogni* $n > 0$*.*

Proof: Dall'identità $\sum_{k=1}^{\infty} f_k = 1 = w_0 = \sum_{k=1}^{\infty} f_k w_k$ ricaviamo che

$$\sum_{k=1}^{\infty} f_k(1 - w_k) = 0$$

che, per la positività dei singoli termini, implica che per ogni $k = 1, 2, \ldots$ risulta o che $f_k = 0$ oppure che $w_k = 1$; in altre parole se $i \in \mathcal{N}$ allora $w_i = 1$. Iterando il ragionamento per ognuno di questi i otteniamo facilmente per induzione che se $m \in \mathcal{N}^+$ allora $w_m = 1$.
Per il lemma precedente esiste un N tale che ogni $m > N$ appartiene a $\mathcal{N}^+$; applicando la (20) otteniamo che $w_N = 1$ e quindi di nuovo per induzione (applicando la (20)) il lemma è dimostrato.

Da fare

- Teorema di Perron-Frobenius
- Matrici di Leslie

Bibliografia

[1] **W. Feller**, An Introduction to Probability Theory and Its Applications, John Wiley & Sons

[2] **A. N. Shiryaev**, Probability, Springer Verlag