

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea in Matematica

Teorema ergodico per
catene di Markov e algoritmo
PageRank di Google

Relatore:
Chiar.mo Prof.
Andrea Cosso

Presentata da:
Sascha Portaro

II Sessione
Anno Accademico 2020/2021

*Ai miei genitori, a mio fratello e a mia sorella,
alla mia famiglia tutta e alle persone più care.
Non avrei mai potuto raggiungere questo
risultato senza il vostro appoggio.*

Indice

Introduzione	5
1 Le catene di Markov	7
1.1 Processi stocastici	7
1.2 Catene di Markov a tempo discreto	9
1.3 Catene di Markov omogenee e a stati finiti	13
1.3.1 Classi comunicanti	17
1.3.2 Legge di una catena di Markov e distribuzione invariante	20
2 Il Teorema ergodico	22
2.1 Catene di Markov regolari	22
2.2 Teorema ergodico	23
3 Applicazione al PageRank	28
3.1 Google e l'algoritmo PageRank	28
3.1.1 Descrizione dell'algoritmo	29
3.2 Esempio di calcolo del PageRank	32
Bibliografia	37

Introduzione

Al giorno d'oggi c'è sempre una maggiore richiesta di strumenti matematici utili ad affrontare vari problemi che si presentano nello sviluppo di tecnologie sempre più sofisticate. È anche per questo motivo che in questa dissertazione trattiamo un importante argomento della teoria della probabilità, ovvero le catene di Markov a tempo discreto e la loro applicazione all'algoritmo PageRank di Google, sviluppato tra il 1996 e il 1998 da Sergey Brin e Larry Page, che permette di quantificare l'importanza di una pagina web e, di conseguenza, stabilire un ordine in cui presentare tali pagine. Ma, al di là della loro utilità pratica, lo studio delle catene di Markov è affascinante in quanto mette in relazione i processi stocastici con altre strutture matematiche, come ad esempio i grafi e le matrici.

Le catene di Markov sono processi stocastici a tempo discreto, la cui evoluzione non è deterministica ma probabilistica ed è caratterizzata dal fatto che la probabilità di passare da uno stato ad un altro non dipende dal percorso fatto fino a quel momento (cioè da tutti gli stati attraversati fino a quel momento), ma dipende esclusivamente dallo stato attuale.

Il cuore di tutta la tesi è il secondo capitolo, dove viene presentato un fondamentale risultato riguardante il comportamento asintotico di una particolare classe di catene di Markov a tempo discreto, ossia il Teorema ergodico. Tale risultato ci consentirà di affermare che, per tempi sufficientemente grandi, la probabilità di raggiungere un certo stato finale non dipende dallo stato di partenza.

Infine concluderemo l'elaborato con un esempio di determinazione del PageRank per una struttura topologica di grafo che descrive un web in miniatura, che ci permetterà di avere un'idea pratica di come funziona l'algoritmo.

Capitolo 1

Le catene di Markov

In questo capitolo sono date per acquisite conoscenze di base della teoria della probabilità che si possono trovare in vari libri tra cui ad esempio [4].

Iniziamo il capitolo fornendo alcune nozioni di base che riguardano i processi stocastici, dopodiché ci concentriamo sulle catene di Markov a tempo discreto, sulle sue caratteristiche, sullo studio di alcune sue proprietà e dimostriamo dei risultati che ci saranno utili per dimostrare il Teorema ergodico che tratteremo nel prossimo capitolo. Trattandosi di argomenti molto basilari, essi sono affrontati in praticamente tutti i libri introduttivi ai processi stocastici, noi abbiamo seguito principalmente [1] e [5, cap. 6].

Nel seguito lavoreremo sempre su uno spazio di probabilità $(\Omega, \mathcal{F}, \mathbb{P})$, dove Ω è lo spazio campionario, \mathcal{F} è la σ -algebra degli eventi e \mathbb{P} è una misura di probabilità.

1.1 Processi stocastici

Preliminarmente, per poter definire le catene di Markov¹, introduciamo la definizione generale di *processo stocastico* (anche detto *processo casuale* o *processo aleatorio*), distinguendo due casi in base alla natura del parametro temporale. Un processo stocastico non è altro che una generalizzazione del concetto di variabile aleatoria e serve a descrivere matematicamente una quantità incerta che evolve nel tempo.

Definizione 1.1.1. Sia $(\Omega, \mathcal{F}, \mathbb{P})$ uno spazio di probabilità.

- Si chiama **processo stocastico a tempo continuo** una famiglia di variabili aleatorie $X = (X_t)_{t \in I}$, con $I = [0, +\infty[$ e $X_t : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})^2, \forall t \in I$.

¹Andrej Andreevič Markov (Rjazan', 14 giugno 1856 – San Pietroburgo, 20 luglio 1922) è stato un matematico e statistico russo che ha dato un grande contributo allo sviluppo della teoria della probabilità.

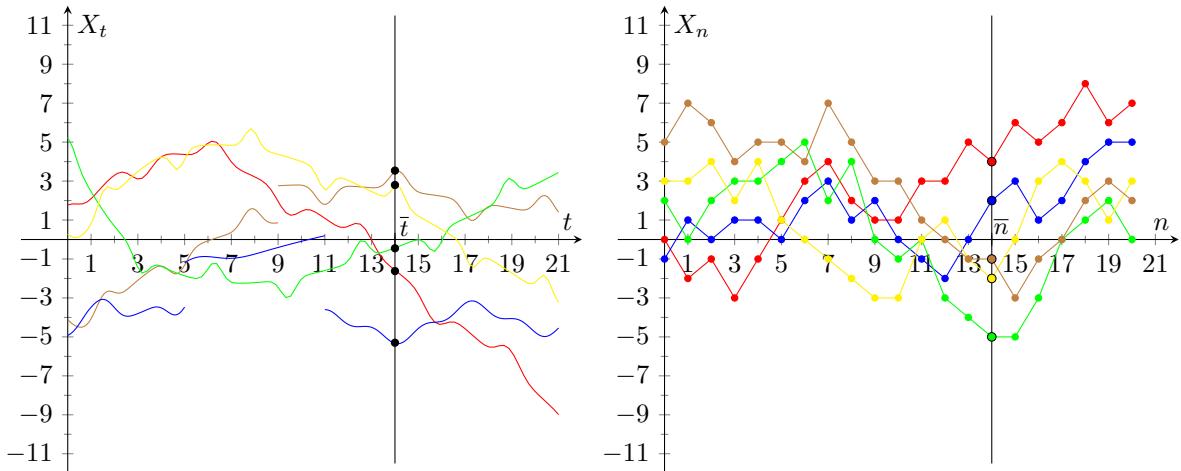
² \mathcal{B} rappresenta la σ -algebra di Borel di \mathbb{R} .

- Si chiama **processo stocastico a tempo discreto** una famiglia di variabili aleatorie $X = (X_n)_{n \in I}$, con $I \subseteq \mathbb{N}$ e $X_n : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}), \forall n \in \mathbb{N}$.

Nel seguito considereremo soltanto processi stocastici a tempo discreto, in cui inoltre $X_1, X_2, \dots, X_n, \dots$ sono tutte variabili aleatorie discrete³ e in particolare con supporto⁴ finito.

Osservazione 1.1.2. Se $X = (X_n)_{n \in I}$, con $I \subseteq \mathbb{N}$, è un processo stocastico a tempo discreto, possiamo vederlo come $X : \Omega \rightarrow \mathbb{R}^I, \Omega \ni \omega \mapsto (X_n)_{n \in I} \in \mathbb{R}^I$, dove $\mathbb{R}^I = \{(x_n)_{n \in I} : x_n \in \mathbb{R}\} = \{\text{Successioni di numeri reali aventi } I \text{ termini}\}$. $\forall \omega \in \Omega$, la funzione $I \ni n \mapsto X_n(\omega)$ si dice una *traiettoria* di X . Una *traiettoria* rappresenta una possibile evoluzione del fenomeno descritto da X nel tempo.

Vediamo ora una rappresentazione grafica di un processo stocastico a tempo continuo e di uno a tempo discreto.



(a) Rappresentazione di un processo stocastico a tempo continuo.

(b) Rappresentazione di un processo stocastico a tempo discreto (le linee che uniscono i punti non fanno parte del processo, sono messe solamente per rendere più comprensibile il grafico).

Figura 1.1: In entrambi i casi, sia per i processi a tempo continuo che per quelli a tempo discreto, colori diversi rappresentano traiettorie diverse. Inoltre notiamo che se fissiamo un tempo $\bar{t} \in [0, +\infty[$ nel primo caso, oppure $\bar{n} \in \mathbb{N}$ nel secondo caso, i punti in cui la retta verticale $t = \bar{t}$ (oppure $n = \bar{n}$ nel caso discreto) incontra le varie traiettorie rappresenta la variabile aleatoria $X_{\bar{t}}$ ($X_{\bar{n}}$ nel caso discreto).

³Una variabile aleatoria $X : \Omega \rightarrow \mathbb{R}$ si dice discreta se esiste un insieme finito o numerabile D tale che $\mathbb{P}(X \in D) = 1$.

⁴Il supporto di una variabile aleatoria discreta X è l'insieme $\mathcal{S}_X := \{x \in \mathbb{R} : \mathbb{P}(X = x) > 0\}$.

1.2 Catene di Markov a tempo discreto

Per poter dare la definizione di catena di Markov sono necessarie alcune definizioni preliminari.

D'ora in avanti I sarà sempre un sottoinsieme di \mathbb{N} ($I \subseteq \mathbb{N}$) e spesso coinciderà con \mathbb{N} stesso.

Definizione 1.2.1. Una famiglia di σ -algebre $(\mathcal{F}_n)_{n \in I}$, con $I \subseteq \mathbb{N}$, si dice una *filtrazione* se $\forall n, m \in \mathbb{N}$, con $n < m$, si ha $\mathcal{F}_n \subseteq \mathcal{F}_m \subseteq \mathcal{F}$.

Definizione 1.2.2. Un processo stocastico a tempo discreto $X = (X_n)_{n \in I}$ si dice *adattato* ad una filtrazione $(\mathcal{F}_n)_{n \in I}$ se $X_n \in m\mathcal{F}_n^5$, $\forall n \in I$.

Osservazione 1.2.3. Osserviamo che una filtrazione $(\mathcal{F}_n)_{n \in I}$ è monotona crescente ed inoltre ha la seguente interpretazione:

- \mathcal{F}_n rappresenta gli eventi osservabili fino al tempo n ;
- tutti gli eventi osservabili al tempo n lo sono anche al tempo m (se $n < m$).

Parallelamente il concetto di processo adattato si può interpretare come processo "osservabile".

D'altra parte, potendo osservare il processo $X = (X_n)_{n \in I}$, potremmo chiederci quali altri eventi sono osservabili. Diamo quindi la definizione di *filtrazione naturale*.

Definizione 1.2.4. Sia $X = (X_n)_{n \in I}$ un processo stocastico a tempo discreto. Si dice *filtrazione naturale* di X la famiglia $(\mathcal{F}_n^X)_{n \in I}$ data da:

$$\mathcal{F}_n^X := \sigma(X_i, i \leq n)^6 = \sigma(\{X_i \in H\}, i \leq n, H \in \mathcal{B}).$$

Osservazione 1.2.5. $(\mathcal{F}_n^X)_{n \in I}$ è la più piccola filtrazione che rende X adattato.

Diamo ora la definizione di proprietà di Markov più generale di quella che useremo nel seguito.

Definizione 1.2.6. Un processo stocastico a tempo discreto $X = (X_n)_{n \in I}$ definito su $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \in I}, \mathbb{P})$, con $(\mathcal{F}_n)_{n \in I}$ filtrazione, ha la *proprietà di Markov* se è adattato e $\forall \varphi \in bm\mathcal{B}^7$, $\forall n \in I$ vale:

$$\mathbb{E}[\varphi(X_{n+1}) | \mathcal{F}_n] = \mathbb{E}[\varphi(X_{n+1}) | X_n]. \quad (1.2.1)$$

⁵ $m\mathcal{F}_n$ indica l'insieme delle variabili aleatorie \mathcal{F}_n -misurabili.

⁶ $\sigma(X_i) := X_i^{-1}(\mathcal{B}) = \{X_i^{-1}(H) | H \in \mathcal{B}\}$ è la σ -algebra generata da X_i .

⁷ $bm\mathcal{B}$ è l'insieme delle funzioni \mathcal{B} -misurabili e limitate.

Osservazione 1.2.7. L'uguaglianza in (1.2.1) è intesa per ogni coppia di versioni di valore atteso condizionato.

Proposizione 1.2.8. Un processo stocastico a tempo discreto $X = (X_n)_{n \in I}$ definito su $(\Omega, \mathcal{F}, \mathbb{P})$ ha la proprietà di Markov rispetto alla filtrazione $(\mathcal{F}_n)_{n \in \mathbb{N}}$ se e solo se vale:

$$\mathbb{E}[\mathbb{1}_{(X_{n+1} \in A)} | \mathcal{F}_n] = \mathbb{E}[\mathbb{1}_{(X_{n+1} \in A)} | X_n], \forall n \in \mathbb{N}, \forall A \in \mathcal{B}. \quad (1.2.2)$$

Dimostrazione. Dobbiamo mostrare che (1.2.1) è equivalente a (1.2.2).

Ovviamente se vale (1.2.1) vale anche (1.2.2), poiché $\mathbb{1}_{(X_{n+1} \in A)} \in \text{bm}\mathcal{B}$, $\forall n \in \mathbb{N}$.

Proviamo quindi l'implicazione inversa.

Supponiamo valga (1.2.2) e fissiamo $n \in \mathbb{N}$ e $\varphi \in \text{bm}\mathcal{B}$.

Sia $(\varphi_m)_{m \in \mathbb{N}}$ una successione di funzioni semplici e limitate tali che $\varphi_m \xrightarrow{m \rightarrow +\infty} \varphi$ puntualmente, supponiamo che sia, $\forall m \in \mathbb{N}$, $\varphi_m = \sum_{j \in J^{(m)}} c_j^{(m)} \mathbb{1}_{A_j^{(m)}}$, dove $J^{(m)} \subset \mathbb{N}$ limitato

e $c_j^{(m)}$ costanti, $A_j^{(m)} \in \mathcal{B}$ per ogni $j \in J^{(m)}$ (una tale successione esiste sempre, per maggiore dettagli si veda [3, pp. 356-357]).

Per il Teorema della convergenza dominata di Lebesgue per il valore atteso condizionato, si hanno le seguenti:

$$\mathbb{E}[\varphi(X_{n+1}) | \mathcal{F}_n] = \lim_{m \rightarrow +\infty} \mathbb{E}[\varphi_m(X_{n+1}) | \mathcal{F}_n];$$

$$\mathbb{E}[\varphi(X_{n+1}) | X_n] = \lim_{m \rightarrow +\infty} \mathbb{E}[\varphi_m(X_{n+1}) | X_n].$$

Inoltre per la (1.2.2) abbiamo:

$$\begin{aligned} \lim_{m \rightarrow +\infty} \mathbb{E}[\varphi_m(X_{n+1}) | \mathcal{F}_n] &= \lim_{m \rightarrow +\infty} \sum_{j \in J^{(m)}} c_j^{(m)} \mathbb{E}[\mathbb{1}_{(X_{n+1} \in A_j^{(m)})} | \mathcal{F}_n] \\ &= \lim_{m \rightarrow +\infty} \sum_{j \in J^{(m)}} c_j^{(m)} \mathbb{E}[\mathbb{1}_{(X_{n+1} \in A_j^{(m)})} | X_n] = \lim_{m \rightarrow +\infty} \mathbb{E}[\varphi_m(X_{n+1}) | X_n]. \end{aligned}$$

E mettendo insieme queste uguaglianze si ottiene (1.2.1). \square

NOTAZIONE. Siano $X = (X_n)_{n \in I}$ un processo stocastico a tempo discreto definito su $(\Omega, \mathcal{F}, \mathbb{P})$, $(\mathcal{F}_n)_{n \in \mathbb{N}}$ una filtrazione e $A \in \mathcal{B}$. Nel seguito useremo la seguente notazione: $\mathbb{P}(X_{n+1} \in A | \mathcal{F}_n) := \mathbb{E}[\mathbb{1}_{(X_{n+1} \in A)} | \mathcal{F}_n]$ e $\mathbb{P}(X_{n+1} \in A | X_n) := \mathbb{E}[\mathbb{1}_{(X_{n+1} \in A)} | X_n]$.

Proposizione 1.2.9. Sia $X = (X_n)_{n \in \mathbb{N}}$ un processo stocastico a tempo discreto su $(\Omega, \mathcal{F}, \mathbb{P})$ con la proprietà di Markov rispetto alla filtrazione $(\mathcal{F}_n)_{n \in \mathbb{N}}$.

Se $(\mathcal{G}_n)_{n \in \mathbb{N}}$ è una filtrazione tale che:

$$(i) \mathcal{G}_n \subseteq \mathcal{F}_n, \forall n \in \mathbb{N};$$

(ii) X è adattato a \mathcal{G}_n .

Allora $X = (X_n)_{n \in \mathbb{N}}$ ha la proprietà di Markov rispetto a $(\mathcal{G}_n)_{n \in \mathbb{N}}$.

In particolare se X ha la proprietà di Markov rispetto ad una qualunque filtrazione, allora la possiede anche rispetto alla filtrazione naturale.

Dimostrazione. Vogliamo provare che vale (1.2.1) per la filtrazione $(\mathcal{G}_n)_{n \in \mathbb{N}}$.

Fissiamo $n \in \mathbb{N}$ e $\varphi \in \text{bm}\mathcal{B}$.

Usando l'ipotesi (i) e la proprietà della torre si ha:

$$\mathbb{E}[\varphi(X_{n+1})|\mathcal{G}_n] = \mathbb{E}[\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n]|\mathcal{G}_n]. \quad (1.2.3)$$

Inoltre X ha la proprietà di Markov rispetto a $(\mathcal{F}_n)_{n \in \mathbb{N}}$, cioè vale (1.2.1) per $(\mathcal{F}_n)_{n \in \mathbb{N}}$:

$$\mathbb{E}[\varphi(X_{n+1})|\mathcal{F}_n] = \mathbb{E}[\varphi(X_{n+1})|X_n].$$

Da (1.2.1) per $(\mathcal{F}_n)_{n \in \mathbb{N}}$ e (1.2.3) si ottiene:

$$\mathbb{E}[\varphi(X_{n+1})|\mathcal{G}_n] = \mathbb{E}[\mathbb{E}[\varphi(X_{n+1})|X_n]|\mathcal{G}_n]. \quad (1.2.4)$$

Dall'ipotesi (ii) si ha che $\mathbb{E}[\varphi(X_{n+1})|X_n] \in m\mathcal{G}_n$ e quindi:

$$\mathbb{E}[\mathbb{E}[\varphi(X_{n+1})|X_n]|\mathcal{G}_n] = \mathbb{E}[\varphi(X_{n+1})|X_n]. \quad (1.2.5)$$

Infine da (1.2.4) e (1.2.5) si ha:

$$\mathbb{E}[\varphi(X_{n+1})|\mathcal{G}_n] = \mathbb{E}[\varphi(X_{n+1})|X_n],$$

ovvero vale la (1.2.1) per $(\mathcal{G}_n)_{n \in \mathbb{N}}$. □

Definizione 1.2.10. *Un processo stocastico a tempo discreto a valori in uno spazio discreto si dice un processo discreto.*

Possiamo finalmente dare la definizione di catena di Markov a tempo discreto.

Definizione 1.2.11. *Un processo discreto si dice una catena di Markov discreta se ha la proprietà di Markov rispetto alla filtrazione naturale.*

Riportiamo ora una caratterizzazione della proprietà di Markov per processi discreti che utilizzeremo nel seguito.

Teorema 1.2.12. *Un processo discreto $X = (X_n)_{n \in \mathbb{N}}$ su $(\Omega, \mathcal{F}, \mathbb{P})$ tale che il supporto di ogni X_n è contenuto in un insieme discreto $\mathcal{S} \subset \mathbb{R}$, è una catena di Markov discreta se e solo se $\forall i_0, \dots, i_n, j \in \mathcal{S}$ tali che $\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$ vale:*

$$\mathbb{P}(X_{n+1} = j | X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j | X_n = i_n), \quad (1.2.6)$$

$\forall n \in \mathbb{N}$.

Dimostrazione. Siano $A_{i_0, \dots, i_n} := \bigcap_{j=0}^n (X_j = i_j)$ e $B_n := (X_n = i_n)$. Vogliamo dimostrare che (1.2.1) è equivalente a:

$$\mathbb{P}(X_{n+1} = j | A_{i_0, \dots, i_n}) = \mathbb{P}(X_{n+1} = j | B_n), \quad \forall n \in \mathbb{N}, \forall i_0, \dots, i_n, j \text{ tali che } \mathbb{P}(A_{i_0, \dots, i_n}) > 0. \quad (1.2.7)$$

$\forall i_0, \dots, i_n, j$ tali che $\mathbb{P}(A_{i_0, \dots, i_n}) > 0$ e $\forall \omega \in A_{i_0, \dots, i_n}$ valgono le seguenti:

$$\mathbb{P}(X_{n+1} = j | X_n)(\omega) = \mathbb{P}(X_{n+1} = j | B_n), \text{ poiché } \omega \in B_n \subset A_{i_0, \dots, i_n}.$$

$$\mathbb{P}(X_{n+1} = j | \mathcal{F}_n^X)(\omega) = \mathbb{P}(X_{n+1} = j | A_{i_0, \dots, i_n}), \text{ poiché } \mathcal{F} = \sigma(X_0, \dots, X_n).$$

Quindi se vale (1.2.1), in virtù della proposizione 1.2.8, vale anche (1.2.7).

Viceversa se vale (1.2.7) vale anche (1.2.1), perché gli eventi A_{i_0, \dots, i_n} formano una partizione numerabile di Ω . \square

NOTAZIONE. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov discreta e assumiamo che $\forall n \in \mathbb{N}$ si ha $\mathcal{S}_{X_n} \subseteq \mathcal{S}$, dove \mathcal{S} è un insieme discreto e con \mathcal{S}_{X_n} abbiamo indicato il supporto di X_n . Allora, per semplicità, nel seguito considereremo solo \mathcal{S} invece di tutti gli \mathcal{S}_{X_n} anche se ciascuna variabile aleatoria non assumerà necessariamente tutti i valori contenuti in \mathcal{S} . Chiameremo \mathcal{S} lo **spazio degli stati** di X .

Osservazione 1.2.13. La proprietà di Markov (1.2.6) fa intuire il perché del nome "catena", infatti vuol dire che le variabili aleatorie di una catena di Markov discreta non sono indipendenti, ma hanno una struttura di dipendenza abbastanza semplice (ognuna dipende solo da quella che la precede).

Osservazione 1.2.14. Se $X = (X_n)_{n \in \mathbb{N}}$ è una catena di Markov discreta, fissiamo un tempo $n \in \mathbb{N}$ e consideriamo gli eventi:

$$A := (X_{n-1} = i_{n-1}) \cap \dots \cap (X_0 = i_0), \text{ con } i_{n-1}, \dots, i_0 \in \mathcal{S};$$

$$B := (X_n = i_n);$$

$$C := (X_{n+1} = j_1) \cap \dots \cap (X_{n+k} = j_k), \text{ con } k \in \mathbb{N}, k > 0, j_1, \dots, j_k \in \mathcal{S}.$$

L'evento A può essere interpretato come il "passato", l'evento B come il "presente" e l'evento C come il "futuro". Allora per la condizione (1.2.6) si ha:

$$\mathbb{P}(C | A \cap B) = \mathbb{P}(C | B). \quad (1.2.8)$$

Da cui, per definizione di probabilità condizionata, otteniamo:

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

$$\stackrel{\substack{= \\ \uparrow \\ (1.2.8)}}{=} \mathbb{P}(C|B) \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(C|B)\mathbb{P}(A|B).$$

Questa uguaglianza si può interpretare come "passato e futuro sono indipendenti condizionatamente al valore presente".

La condizione (1.2.6) del Teorema 1.2.12 è la definizione classica di catena di Markov discreta (come in [3]) e ci permette di dare la definizione seguente.

Definizione 1.2.15. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov discreta. $\forall i_0, \dots, i_n, j \in \mathcal{S}$ tali che $\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0$ la quantità

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

si chiama **probabilità di transizione all'istante n dallo stato i allo stato j** .

1.3 Catene di Markov omogenee e a stati finiti

In questa sezione tratteremo una classe di catene di Markov molto importante per gli argomenti che affronteremo, ovvero le *catene di Markov omogenee e a stati finiti*.

Definizione 1.3.1. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov discreta.

Si dice che X è **omogenea** (nel tempo) se la probabilità di transizione dallo stato i allo stato j non dipende dal tempo n . In tal caso, si scrive π_{ij} anziché $\pi_{ij}(n)$ e si dice che π_{ij} è la **probabilità di transizione dallo stato i allo stato j** .

Definizione 1.3.2. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov discreta.

Si dice che X è **a stati finiti** se il suo spazio degli stati \mathcal{S} è un insieme finito. In tal caso indicheremo con N la cardinalità di \mathcal{S} e per comodità, assumeremo $\mathcal{S} = \{1, \dots, N\}$.

La struttura di dipendenza in una catena di Markov omogenea e a stati finiti è totalmente descritta da una matrice quadrata di ordine N . Pertanto è utile dare la definizione di *matrice di transizione*.

Definizione 1.3.3. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov omogenea e a stati finiti.

Si chiama **matrice di transizione** di X la matrice $N \times N$ (dove N è la cardinalità dello spazio degli stati di X), indicata con Π , le cui componenti sono le probabilità di transizione:

$$\pi_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) \quad \forall i, j \in \mathcal{S},$$

con n qualunque, dato che la catena è omogenea. Esplicitamente: $\Pi = (\pi_{ij})_{i,j \in \mathcal{S}}$.

Osservazione 1.3.4. Dalla definizione di matrice di transizione segue immediatamente che la probabilità di transizione π_{ij} corrisponde all'elemento nella riga i e colonna j della matrice.

Se Π è la matrice di transizione di una catena di Markov omogenea e a stati finiti $X = (X_n)_{n \in \mathbb{N}}$, allora i suoi elementi sono delle probabilità e ogni sua riga i corrisponde alla densità discreta di X_{n+1} condizionata all'evento $(X_n = i)$. Vale quindi il seguente Teorema.

Teorema 1.3.5. *Sia Π la matrice di transizione di una catena di Markov omogenea e a stati finiti $X = (X_n)_{n \in \mathbb{N}}$. Allora Π è tale che:*

- 1) $0 \leq \pi_{ij} \leq 1, \forall i, j \in \mathcal{S}$;
- 2) $\sum_{j=1}^N \pi_{ij} = 1, \forall i \in \mathcal{S}$.

Dimostrazione. $\forall i, j \in \mathcal{S}$ si ha $\pi_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) \in [0, 1]$, cioè vale (1). Gli eventi $(X_{n+1} = j)$, con $j \in \mathcal{S}$, costituiscono una partizione di Ω , quindi per la *formula delle probabilità totali* si ha:

$$\mathbb{P}(X_n = i) = \sum_{j=1}^N \mathbb{P}(X_n = i, X_{n+1} = j). \quad (1.3.1)$$

Inoltre abbiamo che:

$$\sum_{j=1}^N \pi_{ij} = \sum_{j=1}^N \mathbb{P}(X_{n+1} = j | X_n = i) = \sum_{j=1}^N \frac{\mathbb{P}(X_{n+1} = j, X_n = i)}{\mathbb{P}(X_n = i)}.$$

Infine usando la (1.3.1) si ha:

$$\sum_{j=1}^N \frac{\mathbb{P}(X_{n+1} = j, X_n = i)}{\mathbb{P}(X_n = i)} = \frac{1}{\mathbb{P}(X_n = i)} \sum_{j=1}^N \mathbb{P}(X_{n+1} = j, X_n = i) = \frac{\mathbb{P}(X_n = i)}{\mathbb{P}(X_n = i)} = 1.$$

Mettendo assieme tutte le precedenti uguaglianze si ha la (2). □

Data una catena di Markov omogenea e a stati finiti, ci potremmo chiedere quale sia la probabilità di raggiungere lo stato j in m passi partendo dallo stato i . Questa probabilità è definita nel modo seguente.

Definizione 1.3.6. *Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov omogenea e a stati finiti. Per ogni $m \in \mathbb{N}$, poniamo*

$$\pi_{ij}^{(m)} := \mathbb{P}(X_{n+m} = j | X_n = i), \forall i, j \in \mathcal{S} \text{ (con } n \text{ qualsiasi)}.$$

$\pi_{ij}^{(m)}$ si chiama **probabilità di transizione dallo stato i allo stato j in m passi**.

Teorema 1.3.7. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov omogenea e a stati finiti. Se $\forall m \in \mathbb{N}$, $\Pi^{(m)}$ è la matrice di componenti $\pi_{ij}^{(m)}$, allora vale:

$$\Pi^{(m)} = \Pi^m = \underbrace{\Pi \cdots \Pi}_{m \text{ volte}}. \quad (1.3.2)$$

Dimostrazione. Diamo una dimostrazione per induzione su $m \in \mathbb{N}$.

Se $m = 0$, allora $\pi_{ij}^{(0)} = \mathbb{P}(X_n = j | X_n = i)$. Vogliamo dimostrare che $\Pi^{(0)} = \Pi^0 = I_N$, dove I_N è la matrice identità $N \times N$.

Osserviamo che:

$$(X_n = j) \cap (X_n = i) = \begin{cases} (X_n = i), & \text{se } i = j, \\ \emptyset, & \text{se } i \neq j. \end{cases}$$

Quindi, per definizione di probabilità condizionata, si ha:

$$\begin{aligned} \mathbb{P}(X_n = j | X_n = i) &= \frac{\mathbb{P}((X_n = j) \cap (X_n = i))}{\mathbb{P}(X_n = i)} \\ &= \begin{cases} \frac{\mathbb{P}(X_n = i)}{\mathbb{P}(X_n = i)}, & \text{se } i = j, \\ \frac{\mathbb{P}(\emptyset)}{\mathbb{P}(X_n = i)}, & \text{se } i \neq j. \end{cases} = \begin{cases} 1, & \text{se } i = j, \\ 0, & \text{se } i \neq j. \end{cases} \end{aligned}$$

Pertanto si ha la tesi nel caso $m = 0$.

Se $m = 1$, allora la tesi segue direttamente dalla definizione di $\pi_{ij}^{(m)}$ e π_{ij} .

Supponiamo ora che valga (1.3.2) per $m \geq 2$ e dimostriamo che vale anche per $m + 1$.

Per la *formula della probabilità totale*, abbiamo che:

$$\begin{aligned} \pi_{ij}^{(m+1)} &= \mathbb{P}(X_{n+m+1} = j | X_n = i) \\ &= \sum_{i_1, \dots, i_m \in \mathcal{S}} \mathbb{P}(X_{n+m+1} = j, X_{n+m} = i_m, \dots, X_{n+1} = i_1 | X_n = i). \end{aligned}$$

Per definizione di probabilità condizionata, si ha:

$$\begin{aligned} &\sum_{i_1, \dots, i_m \in \mathcal{S}} \mathbb{P}(X_{n+m+1} = j, X_{n+m} = i_m, \dots, X_{n+1} = i_1 | X_n = i) \\ &= \sum_{i_1, \dots, i_m \in \mathcal{S}} \frac{\mathbb{P}(X_{n+m+1} = j, X_{n+m} = i_m, \dots, X_{n+1} = i_1, X_n = i)}{\mathbb{P}(X_n = i)}. \end{aligned}$$

$\forall i_1, \dots, i_m, i, j \in \mathcal{S}$ vale:

$$\frac{\mathbb{P}(X_{n+m+1} = j, X_{n+m} = i_m, \dots, X_{n+1} = i_1, X_n = i)}{\mathbb{P}(X_n = i)}$$

$$\begin{aligned}
&= \frac{\mathbb{P}(X_{n+m} = i_m, \dots, X_{n+1} = i_1, X_n = i)}{\mathbb{P}(X_n = i)} \frac{\mathbb{P}(X_{n+m+1} = j, X_{n+m} = i_m, \dots, X_{n+1} = i_1, X_n = i)}{\mathbb{P}(X_{n+m} = i_m, \dots, X_{n+1} = i_1, X_n = i)} \\
&= \mathbb{P}(X_{n+m} = i_m, \dots, X_{n+1} = i_1 | X_n = i) \mathbb{P}(X_{n+m+1} = j | X_{n+m} = i_m, \dots, X_{n+1} = i_1, X_n = i).
\end{aligned}$$

Per la proprietà di Markov (1.2.6) e per l' omogeneità, abbiamo:

$$\mathbb{P}(X_{n+m+1} = j | X_{n+m} = i_m, \dots, X_{n+1} = i_1, X_n = i) = \mathbb{P}(X_{n+m+1} = j | X_{n+m} = i_m) = \pi_{i_m j}.$$

Quindi, dalle precedenti uguaglianze e usando ancora la *formula della probabilità totale*, si ha:

$$\begin{aligned}
\pi_{ij}^{(m+1)} &= \sum_{i_m \in \mathcal{S}} \left(\sum_{i_1, \dots, i_{m-1} \in \mathcal{S}} \mathbb{P}(X_{n+m} = i_m, \dots, X_{n+1} = i_1 | X_n = i) \right) \pi_{i_m j} \\
&= \sum_{i_m \in \mathcal{S}} \mathbb{P}(X_{n+m} = i_m | X_n = i) \pi_{i_m j} = \sum_{i_m \in \mathcal{S}} \pi_{ii_m}^{(m)} \pi_{i_m j}.
\end{aligned}$$

Per ipotesi induttiva vale $\pi_{ii_m}^{(m)} = \pi_{ii_m}^m$, quindi: $\pi_{ij}^{(m+1)} = \sum_{i_m \in \mathcal{S}} \pi_{ii_m}^m \pi_{i_m j} = \sum_{i_m \in \mathcal{S}} \pi_{ii_m}^m \pi_{i_m j}$, che è esattamente l'elemento nella i -esima riga e j -esima colonna di $\Pi^m \Pi = \Pi^{m+1}$. \square

Rappresentazione grafica di una catena di Markov. Se $X = (X_n)_{n \in \mathbb{N}}$ è una catena di Markov omogenea e a stati finiti, allora si può costruire un *grafo orientato* che la rappresenta nel seguente modo:

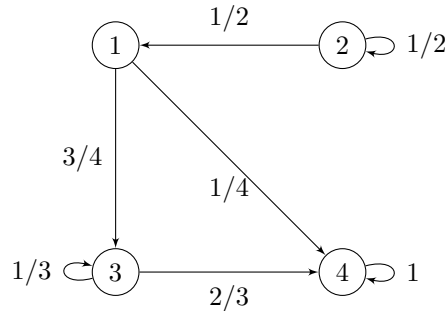
- ogni stato $i \in \mathcal{S}$ corrisponde ad un nodo del grafo;
- se la probabilità di transizione π_{ij} è strettamente positiva si disegna un arco orientato dal nodo i al nodo j , se invece $\pi_{ij} = 0$ non si disegna nessuna freccia dal nodo i al nodo j ;
- su ogni arco orientato si riporta il corrispondente valore di π_{ij} .

Vediamo un esempio di come si rappresenta graficamente una catena di Markov omogenea e a stati finiti.

Esempio 1.3.8. Si consideri una catena di Markov $X = (X_n)_{n \in \mathbb{N}}$ con spazio degli stati $\mathcal{S} = \{1, 2, 3, 4\}$ e matrice di transizione

$$\Pi = \begin{pmatrix} 0 & 0 & 3/4 & 1/4 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Allora il grafo che la rappresenta è il seguente:



NOTAZIONE. Se $X = (X_n)_{n \in \mathbb{N}}$ è una catena di Markov con spazio degli stati \mathcal{S} e $i_0, \dots, i_n \in \mathcal{S}$, un cammino che parte dal nodo i_0 e al k -esimo passo arriva in i_k , con $k = 1, \dots, n$, lo indichiamo con $i_0 \rightarrow i_1 \rightarrow \dots \rightarrow i_n$.

Calcolo diretto della probabilità di transizione in più passi. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov con spazio degli stati \mathcal{S} e matrice di transizione Π . Dal Teorema 1.3.7 sappiamo che $\pi_{ij}^{(m)}$ è l'elemento nella riga i e colonna j della matrice Π^m , dove $m \in \mathbb{N}$ qualsiasi e $i, j \in \mathcal{S}$.

Un modo alternativo e più diretto per determinare $\pi_{ij}^{(m)}$ si basa sull'utilizzo del grafo orientato che rappresenta X e per calcolarlo si procede nel seguente modo:

- si trovano tutti i cammini sul grafo che portano dal nodo i al nodo j in esattamente m passi;
- la probabilità del cammino è il prodotto delle probabilità su ogni arco percorso da tale cammino;
- $\pi_{ij}^{(m)}$ è la somma delle probabilità di tutti i cammini che portano da i a j in m passi.

Calcoliamo $\pi_{2,3}^{(3)}$ dell'esempio 1.3.8.

Ci sono solo due cammini che partono dal nodo 2 e arrivano al nodo 3 in 3 passi, cioè $2 \rightarrow 2 \rightarrow 1 \rightarrow 3$ (la probabilità di questo cammino è $\frac{1}{2} \frac{1}{2} \frac{3}{4} = \frac{3}{16}$) e $2 \rightarrow 1 \rightarrow 3 \rightarrow 3$ (la probabilità di questo cammino è $\frac{1}{2} \frac{3}{4} \frac{1}{3} = \frac{1}{8}$). Pertanto $\pi_{2,3}^{(3)} = \frac{3}{16} + \frac{1}{8} = \frac{5}{16}$.

1.3.1 Classi comunicanti

In questa sottosezione presentiamo una classificazione degli stati che ci permetterà di ottenere una partizione dello spazio degli stati \mathcal{S} di una catena di Markov.

Definizione 1.3.9. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov con spazio degli stati \mathcal{S} . Dati due stati $i, j \in \mathcal{S}$, si dice che j è **accessibile da** i se esiste $m \in \mathbb{N}$ tale che $\pi_{ij}^{(m)} > 0$. In tal caso scriviamo $i \rightsquigarrow j$.

Osservazione 1.3.10. Dati due stati $i, j \in \mathcal{S}$, in termini di grafo orientato, dire che j è accessibile da i significa che il nodo i è **connesso**⁸ con il nodo j .

Osservazione 1.3.11. Poiché $\pi_{ii}^{(0)} = 1$, è sempre vero che i è accessibile da i stesso (in effetti si accede da i a i in $m = 0$ passi). Quindi vale sempre $i \rightsquigarrow i$.

Se $i, j \in \mathcal{S}$ e $i \neq j$, è intuitivamente vero che $i \rightsquigarrow j$ se e solo se esiste un cammino di probabilità positiva (che si può determinare dal grafo) che conduce da i a j . Ciò è quanto afferma il seguente Teorema.

Teorema 1.3.12. Siano $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov con spazio degli stati \mathcal{S} , $i, j \in \mathcal{S}$, con $i \neq j$. Le seguenti affermazioni sono equivalenti:

(i) $i \rightsquigarrow j$;

(ii) esiste $m \in \mathbb{N}$, $m \geq 1$ ed esiste un cammino $i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_{m+1}$ in m passi tale che $i_1 = i$, $i_{m+1} = j$ e $\pi_{i_1 i_2} \pi_{i_2 i_3} \dots \pi_{i_m i_{m+1}} > 0$.

Dimostrazione. Chiamiamo Π la matrice di transizione di X e fissiamo $i, j \in \mathcal{S}$, con $i \neq j$.

Per definizione, la (i) vuol dire che esiste $m \in \mathbb{N}$, $m \geq 1$ tale che $\pi_{ij}^{(m)} > 0$.

Dimostriamo la tesi per induzione su $m \in \mathbb{N}$, $m \geq 1$.

Se $m = 1$ la (i) vuol dire che $\pi_{ij}^{(1)} > 0$, che equivale a $\pi_{ij} > 0$ (per il Teorema 1.3.7), cioè vale (ii) con cammino $i \rightsquigarrow j$.

Sia $m \in \mathbb{N}$, $m \geq 1$. Assumiamo vera la tesi per $m - 1$ e mostriamo che vale anche per m . Per il Teorema 1.3.7 sappiamo che $\pi_{ij}^{(m)}$ è l'elemento nella riga i e colonna j della matrice Π^m . Quindi abbiamo che:

$$\pi_{ij}^{(m)} = \sum_{k=1}^N (\Pi^{m-1})_{ik} \pi_{kj} = \sum_{k=1}^N \pi_{ik}^{(m-1)} \pi_{kj}.$$

Pertanto $\pi_{ij}^{(m)} > 0$ se e solo se almeno un addendo della sommatoria è strettamente positivo (poiché gli addendi sono tutte quantità positive in virtù del Teorema 1.3.5), cioè se e solo se esiste un $k \in \{1, \dots, N\}$ tale che $\pi_{kj} > 0$ e $\pi_{ik}^{(m-1)} > 0$. Dire che $\pi_{ik}^{(m-1)} > 0$ equivale, per ipotesi induttiva, al fatto che esista un cammino $i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_m$ tale che $i_1 = i$, $i_m = k$ e $\pi_{i_1 i_2} \pi_{i_2 i_3} \dots \pi_{i_{m-1} i_m} > 0$.

Dunque vale (i), cioè $\pi_{ij}^{(m)} > 0$, se e solo se il cammino $i_1 \rightarrow i_2 \rightarrow \dots \rightarrow i_m \rightarrow i_{m+1}$ con $i_1 = i$, $i_m = k$ e $i_{m+1} = j$ è tale che $\pi_{i_1 i_2} \pi_{i_2 i_3} \dots \pi_{i_{m-1} i_m} \pi_{i_m i_{m+1}} > 0$, cioè (ii). \square

⁸Dato un grafo orientato e due suoi nodi i e j , si dice che i è *connesso* con j se esiste un cammino da i a j .

Definizione 1.3.13. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov con spazio degli stati \mathcal{S} e fissiamo due stati $i, j \in \mathcal{S}$.

Gli stati i e j si dicono **comunicanti** se $i \rightsquigarrow j$ e $j \rightsquigarrow i$. In tal caso scriviamo $i \longleftrightarrow j$. Un sottoinsieme di \mathcal{S} costituito da tutti gli stati comunicanti tra loro si chiama **classe comunicante**.

Osservazione 1.3.14. In termini di grafo orientato la definizione precedente 1.3.13 si traduce come segue:

- se $i, j \in \mathcal{S}$, $i \longleftrightarrow j$ vuol dire che il nodo i è **fortemente connesso**⁹ con il nodo j ;
- una classe comunicante corrisponde a una **componente fortemente connessa**¹⁰.

Osservazione 1.3.15. Dall'osservazione 1.3.11 è ovvio che $\forall i \in \mathcal{S}$, $i \longleftrightarrow i$, cioè ogni stato è comunicante con se stesso. Ciò implica che ogni stato appartiene ad una classe comunicante (che eventualmente contiene solo tale stato).

Osservazione 1.3.16. Se $X = (X_n)_{n \in \mathbb{N}}$ è una catena di Markov con spazio degli stati \mathcal{S} , dal Teorema 1.3.12 segue che se $i, j \in \mathcal{S}$, $i \neq j$, le seguenti affermazioni sono equivalenti:

- $i \longleftrightarrow j$;
- esiste un cammino chiuso¹¹ che passa per i e per j .

Osservazione 1.3.17. La relazione "essere comunicante con" è una relazione d'equivalenza sullo spazio degli stati \mathcal{S} , cioè soddisfa le seguenti proprietà:

- (1) **riflessività:** $i \longleftrightarrow i$, $\forall i \in \mathcal{S}$;
- (2) **simmetria:** se $i \longleftrightarrow j$ allora $j \longleftrightarrow i$, $\forall i, j \in \mathcal{S}$;
- (3) **transitività:** se $i \longleftrightarrow k$ e $k \longleftrightarrow j$ allora $i \longleftrightarrow j$, $\forall i, j, k \in \mathcal{S}$.

La (1) segue dall'osservazione 1.3.15, la (2) e la (3) invece seguono dall'osservazione 1.3.16.

Pertanto \mathcal{S} può essere partizionato nelle corrispondenti classi d'equivalenza, che sono appunto le classi comunicanti. In particolare ogni stato $i \in \mathcal{S}$ sta in una ed una sola classe comunicante.

Definizione 1.3.18. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov con spazio degli stati \mathcal{S} . Si dice che X è **irriducibile** se esiste un'unica classe comunicante, che è quindi data dall'insieme \mathcal{S} stesso.

⁹Dato un grafo orientato e due suoi nodi i e j , si dice che i è *fortemente connesso* con j se esistono un cammino da i a j e un cammino da j ad i .

¹⁰Una *componente fortemente connessa* di un grafo orientato è un suo sottografo massimale in cui esiste un cammino tra ogni suo nodo.

¹¹Un cammino si dice chiuso se lo stato iniziale coincide con quello finale.

1.3.2 Legge di una catena di Markov e distribuzione invariante

Nelle precedenti sezioni abbiamo descritto alcune proprietà di una catena di Markov omogenea e a stati finiti, adesso cercheremo di dare una descrizione della sua legge, in particolare mostreremo che essa dipende solo dalla distribuzione iniziale e dalla sua matrice di transizione.

Osservazione 1.3.19. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov omogenea e a stati finiti, con spazio degli stati \mathcal{S} . $\forall n \in \mathbb{N}$, per conoscere la densità discreta ν_n di X_n è sufficiente determinare il vettore riga ν_n , dato da:

$$\nu_n := (\nu_n(1) \quad \nu_n(2) \quad \dots \quad \nu_n(N)) = (\mathbb{P}(X_n = 1) \quad \mathbb{P}(X_n = 2) \quad \dots \quad \mathbb{P}(X_n = N)).$$

NOTAZIONE. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov omogenea e a stati finiti. Nel seguito, per qualsiasi $n \in \mathbb{N}$, quando parleremo della legge (o distribuzione) di X_n ci riferiremo al vettore riga ν_n definito come nell'osservazione precedente 1.3.19.

Teorema 1.3.20. Se $X = (X_n)_{n \in \mathbb{N}}$ è una catena di Markov omogenea e a stati finiti con matrice di transizione Π , allora $\forall n \in \mathbb{N}$, la distribuzione di X_n è data dalla seguente formula:

$$\nu_n = \nu_0 \Pi^n.$$

Dimostrazione. Sia \mathcal{S} lo spazio degli stati di X , vogliamo dimostrare che, per ogni $j \in \mathcal{S}$, vale:

$$(\nu_n)_j = \sum_{i=1}^N (\nu_0)_i (\Pi^n)_{ij}.$$

Per definizione, abbiamo le seguenti:

$$(\nu_n)_j = \mathbb{P}(X_n = j).$$

$$(\nu_0)_i = \mathbb{P}(X_0 = i).$$

Inoltre per il Teorema 1.3.7, si ha:

$$(\Pi^n)_{ij} = \pi_{ij}^{(n)} = \mathbb{P}(X_n = j | X_0 = i).$$

Dunque per la *formula della probabilità totale*, abbiamo:

$$\mathbb{P}(X_n = j) = \sum_{i=1}^N \mathbb{P}(X_0 = i) \mathbb{P}(X_n = j | X_0 = i),$$

cioè vale la tesi. □

Definizione 1.3.21. Sia Π la matrice di transizione di una catena di Markov omogenea e a stati finiti $X = (X_n)_{n \in \mathbb{N}}$ e sia $\boldsymbol{\pi}$ un vettore riga di N componenti che verifica le seguenti condizioni:

$$(1) \quad (\boldsymbol{\pi})_i \in [0, 1], \quad \forall i \in \{1, \dots, N\};$$

$$(2) \quad \sum_{i=1}^N (\boldsymbol{\pi})_i = 1.$$

Si dice che $\boldsymbol{\pi}$ è una **distribuzione invariante** (o **stazionaria**, o **di equilibrio**) per Π se vale la formula: $\boldsymbol{\pi} = \boldsymbol{\pi}\Pi$.

Osservazione 1.3.22. Si noti che dire che $\boldsymbol{\pi}$ è una **distribuzione invariante** per una matrice di transizione Π di una catena di Markov omogenea e a stati finiti, equivale a dire che il vettore colonna $\boldsymbol{\pi}^T$ è un autovettore per la matrice Π^T con autovalore 1 e tale che verifichi le condizioni (1) e (2) della definizione precedente 1.3.21.

Teorema 1.3.23. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov omogenea e a stati finiti con matrice di transizione Π . Supponiamo che la distribuzione di X_0 sia invariante, cioè:

$$\boldsymbol{\nu}_0 = \boldsymbol{\nu}_0\Pi. \tag{1.3.3}$$

Allora per ogni $n \in \mathbb{N}$, la distribuzione di X_n è ancora data da $\boldsymbol{\nu}_0$.

Dimostrazione. Diamo una dimostrazione per induzione su $n \in \mathbb{N}$. Se $n = 1$ allora, per il Teorema 1.3.20 e per la (1.3.3), si ha:

$$\boldsymbol{\nu}_1 = \boldsymbol{\nu}_0\Pi = \boldsymbol{\nu}_0.$$

Supponiamo ora che la tesi sia vera per $n \in \mathbb{N}$ e dimostriamo che vale anche per $n + 1$. Sempre per il Teorema 1.3.20, per la (1.3.3) e per l'ipotesi induttiva, abbiamo:

$$\boldsymbol{\nu}_{n+1} = \boldsymbol{\nu}_0\Pi^{n+1} = \boldsymbol{\nu}_0\Pi^n\Pi = \boldsymbol{\nu}_0\Pi = \boldsymbol{\nu}_0.$$

□

Il concetto di distribuzione invariante di una catena di Markov omogenea e a stati finiti è molto importante perché, come vedremo nel prossimo capitolo, è legato al comportamento della catena per lunghi periodi.

Capitolo 2

Il Teorema ergodico

In questo capitolo enunciamo e dimostriamo il Teorema ergodico per le catene di Markov, un risultato notevole che riguarda la convergenza della probabilità di transizione di una catena di Markov regolare e che ci permette quindi di avere un'idea del suo comportamento per tempi sufficientemente grandi.

Per la dimostrazione di questo Teorema ci siamo basati principalmente su [6].

2.1 Catene di Markov regolari

Iniziamo dando la definizione di catena di Markov regolare.

Definizione 2.1.1. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov omogenea e a stati finiti con matrice di transizione Π e spazio degli stati \mathcal{S} . Si dice che X è regolare se esiste $n_0 \in \mathbb{N}$ tale che:

$$\pi_{ij}^{(n_0)} > 0, \quad \forall i, j \in \mathcal{S},$$

ovvero se la matrice $\Pi^{(n_0)}$ ha tutte le componenti positive.

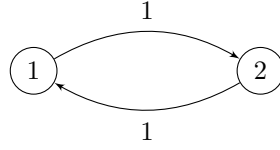
Osservazione 2.1.2. Se una catena di Markov omogenea e a stati finiti è regolare (cioè esiste $n_0 \in \mathbb{N}$ tale che $\pi_{ij}^{(n_0)} > 0$ per ogni i e j nello spazio degli stati), allora si può andare da qualunque stato i ad un qualunque altro stato j al più in n_0 passi. Dunque, in questo caso, X è irriducibile perché esiste un'unica classe comunicante.

Non è vero il viceversa, ossia una catena di Markov omogenea e a stati finiti può essere irriducibile e non regolare. Vediamo un esempio.

Esempio 2.1.3. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov con matrice di transizione

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

e, quindi, grafo orientato



Allora X è irriducibile (perché ogni stato è raggiungibile dall'altro in un solo passo) ma non è regolare. Infatti:

$$\Pi^n = \begin{cases} \Pi, & \text{se } n \text{ è dispari,} \\ I_2, & \text{se } n \text{ è pari,} \end{cases}$$

dove I_2 è la matrice identità 2×2 . Perciò Π^n non ha mai tutte le componenti strettamente positive.

2.2 Teorema ergodico

Teorema 2.2.1 (di convergenza all'equilibrio o ergodico). *Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov omogenea e a stati finiti con matrice di transizione Π e spazio degli stati \mathcal{S} . Se $(X_n)_n$ è regolare allora esiste un'unica distribuzione invariante $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)$ tale che, qualunque sia $i \in \mathcal{S}$,*

$$\lim_{n \rightarrow +\infty} \pi_{ij}^{(n)} = \pi_j, \quad \forall j \in \mathcal{S}. \quad (2.2.1)$$

Inoltre, la velocità di convergenza è esponenziale:

$$|\pi_{ij}^{(n)} - \pi_j| \leq Cq^n, \quad (2.2.2)$$

con $0 \leq q < 1$ e C costante positiva.

Dimostrazione. Poniamo:

$$m_j^{(n)} := \min_{i \in \mathcal{S}} \pi_{ij}^{(n)}, \quad M_j^{(n)} := \max_{i \in \mathcal{S}} \pi_{ij}^{(n)}.$$

Per il Teorema 1.3.7, abbiamo che:

$$\pi_{ij}^{(n+1)} = \sum_{k \in \mathcal{S}} \pi_{ik} \pi_{kj}^{(n)},$$

da cui ricaviamo la seguente disuguaglianza:

$$m_j^{(n+1)} = \min_{i \in \mathcal{S}} \pi_{ij}^{(n+1)} = \min_{i \in \mathcal{S}} \sum_{k \in \mathcal{S}} \pi_{ik} \pi_{kj}^{(n)} \geq \min_{i \in \mathcal{S}} \sum_{k \in \mathcal{S}} \pi_{ik} \min_{h \in \mathcal{S}} \pi_{hj}^{(n)}. \quad (2.2.3)$$

Ora, in virtù del Teorema 1.3.5, si ha $\sum_{k \in \mathcal{S}} \pi_{ik} = 1$, da cui anche $\min_{i \in \mathcal{S}} \sum_{k \in \mathcal{S}} \pi_{ik} = 1$. Pertanto, usando la definizione di $m_j^{(n)}$, dalla (2.2.3) otteniamo:

$$m_j^{(n+1)} \geq m_j^{(n)},$$

cioè la successione $(m_j^{(n)})_{n \in \mathbb{N}}$ è crescente, per ogni $j \in \mathcal{S}$.
Analogamente abbiamo:

$$M_j^{(n+1)} = \max_{i \in \mathcal{S}} \pi_{ij}^{(n+1)} = \max_{i \in \mathcal{S}} \sum_{k \in \mathcal{S}} \pi_{ik} \pi_{kj}^{(n)} \leq \max_{i \in \mathcal{S}} \sum_{k \in \mathcal{S}} \pi_{ik} \max_{h \in \mathcal{S}} \pi_{hj}^{(n)} = M_j^{(n)},$$

cioè la successione $(M_j^{(n)})_{n \in \mathbb{N}}$ è decrescente, per ogni $j \in \mathcal{S}$.
Quindi per provare la (2.2.1), è sufficiente mostrare che

$$M_j^{(n)} - m_j^{(n)} \xrightarrow{n \rightarrow +\infty} 0, \quad \forall j \in \mathcal{S}.$$

Poiché X è regolare, esiste $n_0 \in \mathbb{N}$ tale che: $\pi_{ij}^{(n_0)} > 0, \forall i, j \in \mathcal{S}$. Poniamo $\epsilon := \min_{i, j \in \mathcal{S}} \pi_{ij}^{(n_0)}$ (ovviamente si ha $\epsilon > 0$).

Allora, per il Teorema 1.3.7, abbiamo:

$$\begin{aligned} \pi_{ij}^{(n_0+n)} &= \sum_{k \in \mathcal{S}} \pi_{ik}^{(n_0)} \pi_{kj}^{(n)} = \sum_{k \in \mathcal{S}} \left(\pi_{ik}^{(n_0)} - \epsilon \pi_{jk}^{(n)} \right) \pi_{kj}^{(n)} + \epsilon \sum_{k \in \mathcal{S}} \pi_{jk}^{(n)} \pi_{kj}^{(n)} \\ &= \sum_{k \in \mathcal{S}} \left(\pi_{ik}^{(n_0)} - \epsilon \pi_{jk}^{(n)} \right) \pi_{kj}^{(n)} + \epsilon \pi_{jj}^{(2n)}. \end{aligned} \quad (2.2.4)$$

Ma, poiché $\pi_{jk}^{(n)} \in [0, 1]$, dalla definizione di ϵ si ha: $\pi_{ik}^{(n_0)} - \epsilon \pi_{jk}^{(n)} \geq 0$. Pertanto, da (2.2.4), usando la definizione di $m_j^{(n)}$ e il Teorema 1.3.5, abbiamo:

$$\pi_{ij}^{(n_0+n)} \geq m_j^{(n)} \sum_{k \in \mathcal{S}} \left(\pi_{ik}^{(n_0)} - \epsilon \pi_{jk}^{(n)} \right) + \epsilon \pi_{jj}^{(2n)} = m_j^{(n)} (1 - \epsilon) + \epsilon \pi_{jj}^{(2n)}.$$

Per l'arbitrarietà di i si ottiene:

$$m_j^{(n_0+n)} \geq m_j^{(n)} (1 - \epsilon) + \epsilon \pi_{jj}^{(2n)}. \quad (2.2.5)$$

Analogamente, da (2.2.4), usando la definizione di $M_j^{(n)}$ e il Teorema 1.3.5, si ha:

$$\pi_{ij}^{(n_0+n)} \leq M_j^{(n)} \sum_{k \in \mathcal{S}} \left(\pi_{ik}^{(n_0)} - \epsilon \pi_{jk}^{(n)} \right) + \epsilon \pi_{jj}^{(2n)} = M_j^{(n)} (1 - \epsilon) + \epsilon \pi_{jj}^{(2n)},$$

da cui:

$$M_j^{(n_0+n)} \leq M_j^{(n)} (1 - \epsilon) + \epsilon \pi_{jj}^{(2n)}. \quad (2.2.6)$$

Combinando la (2.2.5) e la (2.2.6), otteniamo:

$$M_j^{(n_0+n)} - m_j^{(n_0+n)} \leq \left(M_j^{(n)} - m_j^{(n)} \right) (1 - \epsilon)$$

e, di conseguenza:

$$\begin{aligned}
M_j^{(hn_0+n)} - m_j^{(hn_0+n)} &\leq \left(M_j^{((h-1)n_0+n)} - m_j^{((h-1)n_0+n)} \right) (1 - \epsilon) \\
&\leq \left(M_j^{((h-2)n_0+n)} - m_j^{((h-2)n_0+n)} \right) (1 - \epsilon)(1 - \epsilon) = \left(M_j^{((h-2)n_0+n)} - m_j^{((h-2)n_0+n)} \right) (1 - \epsilon)^2 \\
&\leq \dots \leq \left(M_j^{(n_0+n)} - m_j^{(n_0+n)} \right) (1 - \epsilon)^{h-1} \leq \left(M_j^{(n)} - m_j^{(n)} \right) (1 - \epsilon)^h. \quad (2.2.7)
\end{aligned}$$

Ma, poiché $0 < \epsilon \leq 1$, si ha che $(1 - \epsilon)^h \xrightarrow{h \rightarrow +\infty} 0$, da cui:

$$M_j^{(kn_0+n)} - m_j^{(kn_0+n)} \xrightarrow{k \rightarrow +\infty} 0.$$

Quindi abbiamo trovato $(M_j^{(n_h)} - m_j^{(n_h)})_{h \in \mathbb{N}}$ sottosuccessione di $(M_j^{(n)} - m_j^{(n)})_{n \in \mathbb{N}}$ che converge a 0 per $h \rightarrow +\infty$. Ma la successione $(M_j^{(n)} - m_j^{(n)})_{n \in \mathbb{N}}$ è monotona decrescente, quindi ha limite per $n \rightarrow +\infty$ e tale limite coincide con quello di ogni sua sottosuccessione, pertanto:

$$M_j^{(n)} - m_j^{(n)} \xrightarrow{n \rightarrow +\infty} 0, \quad \forall j \in \mathcal{S}.$$

Poniamo $\pi_j = \lim_{n \rightarrow +\infty} m_j^{(n)}$, $\forall j \in \mathcal{S}$.

Poiché $(m_j^{(n)})_{n \in \mathbb{N}}$ è monotona crescente, abbiamo:

$$\pi_j = \lim_{n \rightarrow +\infty} m_j^{(n)} \geq m_j^{(n)} \quad \forall j \in \mathcal{S}. \quad (2.2.8)$$

Inoltre, poiché $(M_j^{(n)} - m_j^{(n)})_{n \in \mathbb{N}}$ è monotona decrescente, dalla (2.2.7) e ponendo¹ $h = \left\lceil \frac{n}{n_0} \right\rceil$ (in particolare si ha $hn_0 \leq n$), abbiamo:

$$M_j^{(n)} - m_j^{(n)} \leq M_j^{(hn_0)} - m_j^{(hn_0)} \leq \left(M_j^{(0)} - m_j^{(0)} \right) (1 - \epsilon)^h, \quad \forall j \in \mathcal{S}. \quad (2.2.9)$$

Ma, per il Teorema 1.3.7, si ha:

$$M_j^{(0)} = 1, \quad m_j^{(0)} = 0,$$

quindi dalla (2.2.9) otteniamo:

$$M_j^{(n)} - m_j^{(n)} \leq (1 - \epsilon)^h, \quad \forall j \in \mathcal{S}. \quad (2.2.10)$$

¹Se $x \in \mathbb{R}$ allora con $[x]$ indichiamo la parte intera di x , ovvero $[x] := \max\{n \in \mathbb{Z} \mid n \leq x\}$.

Allora, da (2.2.8), da (2.2.10) e dalla definizione di $M_j^{(n)}$, segue:

$$|\pi_{ij}^{(n)} - \pi_j| \leq M_j^{(n)} - m_j^{(n)} \leq (1 - \epsilon)^{\lfloor \frac{n}{n_0} \rfloor}, \quad \forall j \in \mathcal{S}.$$

Se $\epsilon = 1$ allora $|\pi_{ij}^{(n)} - \pi_j| = 0$, cioè $\pi_{ij}^{(n)} = \pi_j$, non c'è niente da provare. Supponiamo quindi $\epsilon \neq 1$.

Per $n \geq n_0$ dividendo n per n_0 abbiamo $n = a_n n_0 + b_n$, con $a_n, b_n \in \mathbb{N}$, $0 \leq b_n < n_0$, quindi:

$$\left\lfloor \frac{n}{n_0} \right\rfloor = a_n = \frac{n - b_n}{n_0} = n \frac{1}{n_0} - \frac{b_n}{n_0} \geq n \frac{1}{n_0} - \frac{n_0 - 1}{n_0}.$$

Quindi, ponendo $q := (1 - \epsilon)^{\frac{1}{n_0}} \in]0, 1[$ e $C := (1 - \epsilon)^{-\frac{n_0 - 1}{n_0}} > 0$, abbiamo:

$$|\pi_{ij}^{(n)} - \pi_j| \leq Cq^n, \quad \forall j \in \mathcal{S},$$

cioè, qualunque sia $i \in \mathcal{S}$,

$$\lim_{n \rightarrow +\infty} \pi_{ij}^{(n)} = \pi_j, \quad \forall j \in \mathcal{S} \quad (2.2.11)$$

e la velocità di convergenza è esponenziale.

Dobbiamo far vedere ora che $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)$ è una distribuzione invariante.

Dato che $m_j^{(n)} \in [0, 1]$ allora anche $\pi_j \in [0, 1]$, per ogni $j \in \mathcal{S}$. Inoltre, usando la (2.2.11), abbiamo:

$$1 = \sum_{j \in \mathcal{S}} \pi_{ij}^{(n)} \xrightarrow{n \rightarrow +\infty} \sum_{j \in \mathcal{S}} \pi_j,$$

quindi, per l'unicità del limite, otteniamo:

$$\sum_{j \in \mathcal{S}} \pi_j = 1.$$

Infine dobbiamo provare che $\boldsymbol{\pi} = \boldsymbol{\pi}\Pi$, cioè:

$$\pi_j = \sum_{k \in \mathcal{S}} \pi_k \pi_{kj}, \quad \forall j \in \mathcal{S}.$$

Ma, per il Teorema 1.3.7, abbiamo che:

$$\pi_{ij}^{(n+1)} = \sum_{k \in \mathcal{S}} \pi_{ik}^{(n)} \pi_{kj},$$

da cui, usando (2.2.11), si ha:

$$\pi_j \xleftarrow{n \rightarrow +\infty} \pi_{ij}^{(n+1)} = \sum_{k \in \mathcal{S}} \pi_{ik}^{(n)} \pi_{kj} \xrightarrow{n \rightarrow +\infty} \sum_{k \in \mathcal{S}} \pi_k \pi_{kj}, \quad \forall j \in \mathcal{S},$$

cioè, per l'unicità del limite, otteniamo quanto volevamo dimostrare. \square

Osservazione 2.2.2. Sia $X = (X_n)_{n \in \mathbb{N}}$ una catena di Markov omogenea, a stati finiti e regolare con spazio degli stati \mathcal{S} . Il Teorema precedente 2.2.1 fornisce un collegamento tra π_j e la probabilità di transizione $\pi_{ij}^{(n)}$, con $i, j \in \mathcal{S}$; più precisamente afferma che π_j è approssimativamente uguale a $\pi_{ij}^{(n)}$, per n sufficientemente grande. Dunque, per tempi n sufficientemente grandi, π_j rappresenta all'incirca la probabilità di essere allo stato j al tempo n , indipendentemente da quale sia lo stato di partenza i .

Di conseguenza, gli stati j per i quali la probabilità π_j è elevata sono quelli più facilmente raggiungibili dagli altri stati. Quindi, se immaginiamo di percorrere una passeggiata aleatoria lungo il grafo orientato associato a X , gli stati j con probabilità π_j più elevata sono quelli che vengono "visitati" più spesso.

Pertanto, ordinando gli stati in modo crescente in base ai valori contenuti in $\boldsymbol{\pi}$, otteniamo un ordinamento degli stati in base a quanto siano più o meno facilmente "visitabili".

Capitolo 3

Applicazione al PageRank

In questo capitolo vediamo un'applicazione delle catene di Markov all'algoritmo PageRank utilizzato dal motore di ricerca Google, ideato nel 1997 da Sergey Brin e Lawrence "Larry" Page. Per la precisione descriviamo la sua prima versione, in quanto l'algoritmo nella sua forma attuale non è chiaramente noto (essendo ovviamente tutelato da copyright).

Quanto riportato nel seguito si basa su [2].

3.1 Google e l'algoritmo PageRank

Come spiegato in [2], il nome Google è stato scelto da Brin e Page in quanto rimanda al nome *googol*, che è il termine matematico con cui si indica il numero 10^{100} : il motivo è che questo numero rende l'idea della scala dei problemi che un motore di ricerca deve affrontare.

Google, e in particolare l'algoritmo PageRank, si basa sulla struttura "topologica" del web, secondo cui si rappresenta il web come un grafo orientato in cui ogni nodo corrisponde ad una pagina e ogni freccia identifica un link. Tale struttura permette di ordinare le pagine, ossia creare un ranking delle pagine che è appunto il compito dell'algoritmo PageRank. Questo ordinamento viene utilizzato per rispondere in maniera rapida e soddisfacente a ciascuna singola *query*.

L'ordinamento delle pagine web fornito dall'algoritmo PageRank si basa sull'assegnazione di un *indice di significatività* a ciascuna pagina. Per una generica pagina web A , denotiamo tale indice con $PR(A)$. La procedura di assegnazione dell'indice di significatività prescinde dall'effettivo contenuto della pagina, mentre tiene conto di quanto e da chi è citata (ovvero, di quanti e quali link conducono a quella pagina). In particolare, l'indice di significatività deve soddisfare i due requisiti seguenti:

- (1) risultare elevato se una pagina è citata da molte altre pagine;

- (2) risultare elevato se riferito ad una pagina citata da (eventualmente poche) pagine molto significative.

Dunque il solo conteggio dei link ad una pagina non può essere un buon indice di significatività (in quanto non soddisfa il secondo requisito). L'indice di significatività di una pagina A deve essere invece proporzionale agli indici di significatività delle pagine che conducono ad A .

3.1.1 Descrizione dell'algoritmo

Avendo dato al web una struttura topologica come spiegato in precedenza, possiamo immaginare di effettuare una passeggiata aleatoria nel web, scegliendo ad ogni passo un link a caso dalla pagina in cui ci troviamo (più precisamente, supponiamo che i link uscenti da una data pagina siano tra loro equiprobabili). Indichiamo infine con $X = (X_n)_{n \in \mathbb{N}}$ la catena di Markov che descrive tale passeggiata aleatoria e con Π la sua matrice di transizione.

Nella realtà la dimensione della matrice Π è gigantesca, infatti già nel 1997 si stimava la presenza di circa 100 milioni di pagine web e al giorno d'oggi tale stima è arrivata a circa 3,5 miliardi.

Notiamo che gli elementi non nulli della i -esima riga di Π corrispondono alle pagine linkate dalla pagina i . Se tali elementi sono in numero pari a m_i , per l'ipotesi di equiprobabilità hanno tutti lo stesso valore dato da $1/m_i$. Si noti inoltre che gli elementi non nulli della j -esima colonna sono le pagine che hanno un link che conduce alla pagina j . L'idea fondamentale su cui si basa l'algoritmo è quella di considerare la *distribuzione invariante* π della catena di Markov, ordinando gli stati utilizzando proprio l'ordinamento suggerito da π , quindi ponendo

$$\text{PR}(j) := \pi_j$$

per ogni pagina web j .

Vediamo ora perché π_j è un buon indice di significatività, ovvero che rispetta le condizioni (1) e (2).

Data

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} & \dots & \pi_{1N} \\ \pi_{21} & \pi_{22} & \dots & \pi_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{N1} & \pi_{N2} & \dots & \pi_{NN} \end{pmatrix},$$

dobbiamo determinare π tale che $\pi = \pi\Pi$.

Dobbiamo quindi risolvere il seguente sistema:

$$\begin{cases} \pi_1 &= \sum_{i=1}^N \pi_i \pi_{i1} \\ \pi_2 &= \sum_{i=1}^N \pi_i \pi_{i2} \\ \vdots & \\ \pi_N &= \sum_{i=1}^N \pi_i \pi_{iN} \end{cases} \quad (3.1.1)$$

Ricordiamo che $\boldsymbol{\pi}$ è un vettore densità discreta, quindi deve soddisfare le condizioni:

- 1) $0 \leq \pi_j \leq 1$, per ogni $j \in \{1, \dots, N\}$;
- 2) $\sum_{j=1}^N \pi_j = 1$.

Quindi il sistema (3.1.1) ha un'unica soluzione e, per definizione di $\text{PR}(j)$, possiamo riscriverlo in maniera equivalente nel seguente modo:

$$\text{PR}(j) = \sum_{i=1}^N \text{PR}(i) \pi_{ij}, \quad j \in \{1, \dots, N\}.$$

Inoltre, come osservato in precedenza, per l'ipotesi di equiprobabilità si ha che $\frac{1}{\pi_{ij}}$ indica il numero di link che partono dalla pagina i .

Quindi abbiamo che l'indice di significatività di una pagina web A , a cui si accede dalle pagine T_1, \dots, T_n , è dato da

$$\text{PR}(A) = \frac{1}{C(T_1)} \text{PR}(T_1) + \dots + \frac{1}{C(T_n)} \text{PR}(T_n),$$

dove $C(T_i)$ è il numero di link che partono dalla pagina T_i . In conclusione possiamo osservare che l'indice di significatività di A è proporzionale agli indici di significatività delle pagine che conducono ad A , pertanto soddisfa le condizioni (1) e (2) richieste.

Osservazione 3.1.1. *È importante sottolineare che quanto detto finora si basa sul Teorema 2.2.1, che richiede che la catena di Markov X sia regolare. In generale, non è assolutamente garantito che questa ipotesi valga. Per tale ragione vanno presi degli accorgimenti, modificando opportunamente la matrice di transizione Π del web.*

Vediamo quindi come modificare la matrice di transizione Π in modo che diventi regolare.

Innanzitutto osserviamo che nel web ci sono delle pagine da cui non si può accedere ad alcuna altra pagina. Una pagina web di questo tipo si chiama **pagina dangling** e la riga corrispondente sulla matrice di transizione è costituita solo da zeri, dunque Π in tal caso non è neppure una vera matrice di transizione dato che la somma degli elementi di una riga dovrebbe sempre essere uguale a 1.

Dal momento che si può passare da una pagina all'altra non solo tramite i link, ma anche con la barra degli indirizzi, per risolvere il problema delle pagine dangling possiamo ipotizzare che ciascuna di queste punti ad ogni altra pagina del web: questo equivale a sostituire alla riga di una pagina dangling, formata da tutti zeri, il vettore riga

$$\left[\frac{1}{N} \quad \frac{1}{N} \quad \cdots \quad \frac{1}{N} \right].$$

Indichiamo con $\tilde{\Pi}$ la nuova matrice così ottenuta a partire da Π . Si noti che N è talmente grande che $1/N$ è pressoché zero, quindi la matrice $\tilde{\Pi}$ è quasi uguale a Π . Tuttavia, con questa modifica la nuova matrice $\tilde{\Pi}$ è una *vera* matrice di transizione.

Nonostante questa modifica, per poter applicare il Teorema ergodico, la catena di Markov associata a $\tilde{\Pi}$ deve essere *regolare*, ma in generale non è detto che lo sia. Per ottenere tale proprietà è sufficiente perturbare la matrice $\tilde{\Pi}$ nel seguente modo:

$$\Pi_{\text{PR}} = (1 - d) \frac{1}{N} \mathbb{1}_N + d \tilde{\Pi},$$

dove:

- $\mathbb{1}_N$ è la matrice $N \times N$ le cui componenti sono tutte uguali a 1;
- $d \in (0, 1)$ è un parametro fissato (che va scelto in modo “ottimale”).

Si noti che $\frac{1}{N} \mathbb{1}_N$ è una matrice di transizione come $\tilde{\Pi}$, quindi anche la matrice Π_{PR} lo è. Inoltre, dato che tutti gli elementi di Π_{PR} sono strettamente positivi, la catena di Markov associata a Π_{PR} è *regolare*. Possiamo dunque applicare il Teorema 2.2.1.

Il parametro d , detto *damping factor*, è un fattore deciso da Google e nella documentazione originale assume valore 0,85. Può essere aggiustato da Google per decidere la percentuale di PageRank che deve transitare da una pagina all'altra e il valore di PageRank minimo attribuito ad ogni pagina in archivio.

Inoltre possiamo supporre che un navigatore immaginario che percorre una passeggiata aleatoria nel web ad un certo punto smetta di navigare, quindi si può interpretare il *damping factor* come la probabilità che, ad ogni passo, la persona continui a navigare nel web. Ci sono stati vari studi per stimare tale probabilità e generalmente si presume che si aggiri attorno all'85%.

Pertanto possiamo concludere che l'indice di significatività di una pagina web A , a cui si accede dalle pagine T_1, \dots, T_n , è dato da

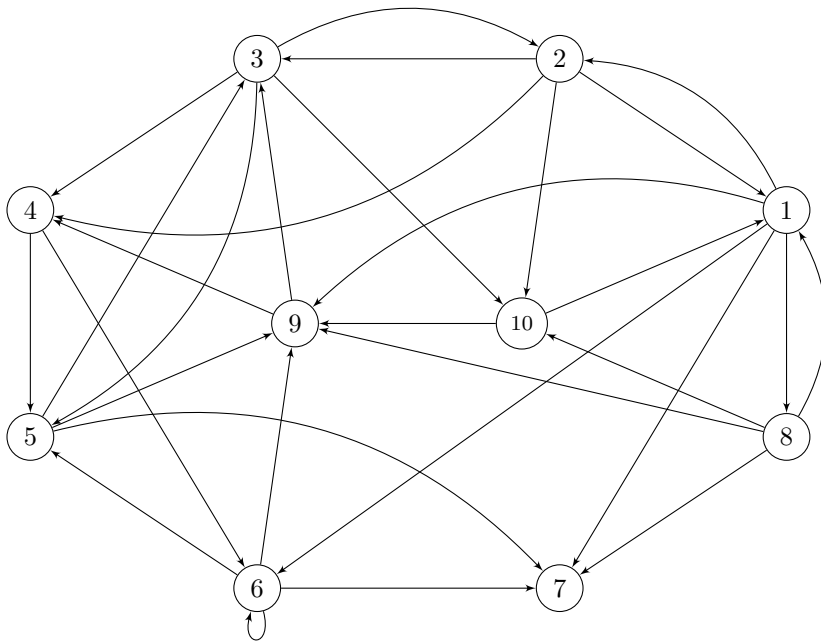
$$\text{PR}(A) = \frac{1 - d}{N} + d \left(\frac{1}{C(T_1)} \text{PR}(T_1) + \cdots + \frac{1}{C(T_n)} \text{PR}(T_n) \right), \quad (3.1.2)$$

con i $C(T_i)$ definiti come prima.

3.2 Esempio di calcolo del PageRank

In questa sezione presentiamo un esempio di calcolo del PageRank in modo da avere un'idea più chiara di come funziona l'algoritmo.

Esempio 3.2.1. *Si consideri una versione semplificata del web, descritta dal seguente grafo orientato, in cui ogni nodo corrisponde ad una pagina, mentre le frecce rappresentano i link tra le pagine:*



Supponiamo di partire dalla pagina numero 1 e di effettuare una passeggiata aleatoria nel web, scegliendo ad ogni passo un link a caso dalla pagina in cui ci troviamo (più precisamente, si suppongano equiprobabili tra loro i link uscenti da una data pagina). Sia $X = (X_n)_n$ la catena di Markov che descrive tale passeggiata aleatoria.

Si chiede di determinare il PageRank per questa versione del web, ovvero assegnare un indice di significatività alle pagine $1, \dots, 10$ che soddisfi i requisiti (1) e (2), in modo da avere un ordinamento di tali pagine.

Come spiegato in precedenza, dobbiamo determinare $\boldsymbol{\pi}$ tale che $\boldsymbol{\pi} = \boldsymbol{\pi}\Pi$, dove

$$\Pi = \begin{pmatrix} 0 & \frac{1}{5} & 0 & 0 & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

Osserviamo che abbiamo utilizzato l'ipotesi di equiprobabilità, inoltre notiamo che la pagina 7 è una pagina dangling perché non c'è nessuna freccia che esce da essa. Pertanto, come abbiamo detto in precedenza, per far diventare la matrice Π una matrice di transizione, sostituiamo alla settima riga, formata da tutti zeri, il vettore riga

$$\left[\frac{1}{10} \quad \frac{1}{10} \quad \frac{1}{10} \quad \frac{1}{10} \quad \frac{1}{10} \quad \frac{1}{10} \quad \frac{1}{10} \quad \frac{1}{10} \quad \frac{1}{10} \quad \frac{1}{10} \right]$$

e otteniamo la matrice

$$\tilde{\Pi} = \begin{pmatrix} 0 & \frac{1}{5} & 0 & 0 & 0 & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} & \frac{1}{10} \\ \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

Ora per far diventare $\tilde{\Pi}$ una matrice regolare, dobbiamo perturbarla come segue:

$$\Pi_{\text{PR}} = (1 - d) \frac{1}{10} \mathbb{1}_{10} + d \tilde{\Pi},$$

dove $d \in (0, 1)$ e $\mathbb{1}_{10}$ sono definiti come prima. In questo modo otteniamo la matrice

$$\Pi_{\text{PR}} = \begin{pmatrix} \frac{1-d}{10} & \frac{1+d}{10} & \frac{1-d}{2+3d} & \frac{1-d}{2+3d} & \frac{1-d}{10} & \frac{1+d}{10} & \frac{1+d}{10} & \frac{1+d}{10} & \frac{1+d}{10} & \frac{1-d}{2+3d} \\ \frac{10}{2+3d} & \frac{10}{1-d} & \frac{10}{2+3d} & \frac{10}{2+3d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{2+3d} \\ \frac{20}{1-d} & \frac{10}{2+3d} & \frac{20}{1-d} & \frac{20}{2+3d} & \frac{20}{2+3d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{20}{2+3d} \\ \frac{10}{1-d} & \frac{20}{1-d} & \frac{10}{1-d} & \frac{20}{1-d} & \frac{20}{1+4d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{20}{1-d} \\ \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{3+7d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{3+7d} & \frac{10}{1-d} & \frac{10}{3+7d} & \frac{10}{1-d} \\ \frac{10}{1-d} & \frac{10}{1-d} & \frac{30}{1-d} & \frac{10}{1-d} & \frac{10}{2+3d} & \frac{10}{2+3d} & \frac{30}{2+3d} & \frac{10}{1-d} & \frac{30}{2+3d} & \frac{10}{1-d} \\ \frac{10}{1} & \frac{10}{1} & \frac{10}{1} & \frac{10}{1} & \frac{20}{1} & \frac{20}{1} & \frac{20}{1} & \frac{10}{1} & \frac{20}{1} & \frac{10}{1} \\ \frac{10}{2+3d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{2+3d} & \frac{10}{1-d} & \frac{10}{2+3d} & \frac{10}{2+3d} \\ \frac{20}{1-d} & \frac{10}{1-d} & \frac{10}{1+4d} & \frac{10}{1+4d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{20}{1-d} & \frac{10}{1-d} & \frac{20}{1-d} & \frac{10}{1-d} \\ \frac{10}{1+4d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1-d} & \frac{10}{1+4d} & \frac{10}{1-d} \end{pmatrix}.$$

Ricordando che $\boldsymbol{\pi}$ deve soddisfare le condizioni per essere un vettore densità discreta, dobbiamo risolvere $\boldsymbol{\pi} = \boldsymbol{\pi}\Pi_{\text{PR}}$, che equivale a $\boldsymbol{\pi}^T = \Pi_{\text{PR}}^T \boldsymbol{\pi}^T$. Cioè dobbiamo trovare gli autovettori di Π_{PR}^T di autovalore 1, la cui esistenza ci è garantita appunto dal Teorema 2.2.1.

Per $d = 0,85$ la matrice di transizione diventa

$$\Pi_{\text{PR}} = \begin{pmatrix} \frac{3}{200} & \frac{37}{200} & \frac{3}{200} & \frac{3}{200} & \frac{3}{200} & \frac{37}{200} & \frac{37}{200} & \frac{37}{200} & \frac{37}{200} & \frac{3}{200} \\ \frac{400}{91} & \frac{200}{3} & \frac{400}{91} & \frac{400}{91} & \frac{200}{3} & \frac{200}{3} & \frac{200}{3} & \frac{200}{3} & \frac{200}{3} & \frac{400}{91} \\ \frac{400}{3} & \frac{200}{91} & \frac{400}{3} & \frac{400}{91} & \frac{200}{91} & \frac{200}{3} & \frac{200}{3} & \frac{200}{3} & \frac{200}{3} & \frac{400}{91} \\ \frac{200}{3} & \frac{400}{3} & \frac{200}{3} & \frac{400}{3} & \frac{400}{11} & \frac{200}{3} & \frac{200}{3} & \frac{200}{3} & \frac{200}{3} & \frac{400}{3} \\ \frac{200}{3} & \frac{200}{3} & \frac{200}{179} & \frac{200}{3} & \frac{25}{3} & \frac{25}{3} & \frac{200}{179} & \frac{200}{3} & \frac{200}{179} & \frac{200}{3} \\ \frac{200}{3} & \frac{200}{3} & \frac{600}{3} & \frac{200}{3} & \frac{200}{91} & \frac{200}{91} & \frac{600}{91} & \frac{200}{3} & \frac{600}{91} & \frac{200}{3} \\ \frac{200}{1} & \frac{200}{1} & \frac{200}{1} & \frac{200}{1} & \frac{400}{1} & \frac{400}{1} & \frac{400}{1} & \frac{200}{1} & \frac{400}{1} & \frac{200}{1} \\ \frac{10}{91} & \frac{10}{3} & \frac{10}{3} & \frac{10}{3} & \frac{10}{3} & \frac{10}{3} & \frac{10}{91} & \frac{10}{3} & \frac{10}{91} & \frac{10}{91} \\ \frac{400}{3} & \frac{200}{3} & \frac{200}{11} & \frac{200}{11} & \frac{200}{3} & \frac{200}{3} & \frac{400}{3} & \frac{200}{3} & \frac{400}{3} & \frac{400}{3} \\ \frac{200}{11} & \frac{200}{3} & \frac{25}{3} & \frac{25}{3} & \frac{200}{3} & \frac{200}{3} & \frac{200}{3} & \frac{200}{3} & \frac{200}{11} & \frac{200}{3} \\ \frac{25}{25} & \frac{200}{200} & \frac{200}{200} & \frac{200}{200} & \frac{200}{200} & \frac{200}{200} & \frac{200}{200} & \frac{200}{200} & \frac{25}{25} & \frac{200}{200} \end{pmatrix}.$$

e la distribuzione invariante $\boldsymbol{\pi}$ è data da:

$$\left\{ \begin{array}{l} \pi_1 \simeq 0,077 \\ \pi_2 \simeq 0,065 \\ \pi_3 \simeq 0,133 \\ \pi_4 \simeq 0,125 \\ \pi_5 \simeq 0,13 \\ \pi_6 \simeq 0,115 \\ \pi_7 \simeq 0,106 \\ \pi_8 \simeq 0,037 \\ \pi_9 \simeq 0,138 \\ \pi_{10} \simeq 0,074 \end{array} \right.$$

Di conseguenza possiamo dire che l'ordinamento delle nostre pagine web è dato da:

$$\text{PR}(9) > \text{PR}(3) > \text{PR}(5) > \text{PR}(4) > \text{PR}(6) > \text{PR}(7) > \text{PR}(1) > \text{PR}(10) > \text{PR}(2) > \text{PR}(8).$$

In generale non serve svolgere tutti i passaggi descritti ma possiamo ottenere lo stesso risultato utilizzando direttamente la formula (3.1.2).

Ad esempio, per calcolare il PageRank della pagina 9, si opera nel seguente modo:

- 1) troviamo le pagine che puntano alla pagina 9, ovvero le pagine 1, 5, 6, 8 e 10;
- 2) contiamo i link uscenti dalle pagine 1, 5, 6, 8 e 10, quindi abbiamo $C(1) = 4$, $C(5) = 3$, $C(6) = 4$, $C(8) = 4$ e $C(10) = 2$.

Allora, utilizzando la (3.1.2), abbiamo:

$$\begin{aligned} \text{PR}(9) &= \frac{1-d}{10} + d \left(\frac{1}{C(1)}\text{PR}(1) + \frac{1}{C(5)}\text{PR}(5) + \frac{1}{C(6)}\text{PR}(6) + \frac{1}{C(8)}\text{PR}(8) + \frac{1}{C(10)}\text{PR}(10) \right) \\ &= \frac{1-d}{10} + d \left(\frac{1}{4}\text{PR}(1) + \frac{1}{3}\text{PR}(5) + \frac{1}{4}\text{PR}(6) + \frac{1}{4}\text{PR}(8) + \frac{1}{2}\text{PR}(10) \right). \end{aligned}$$

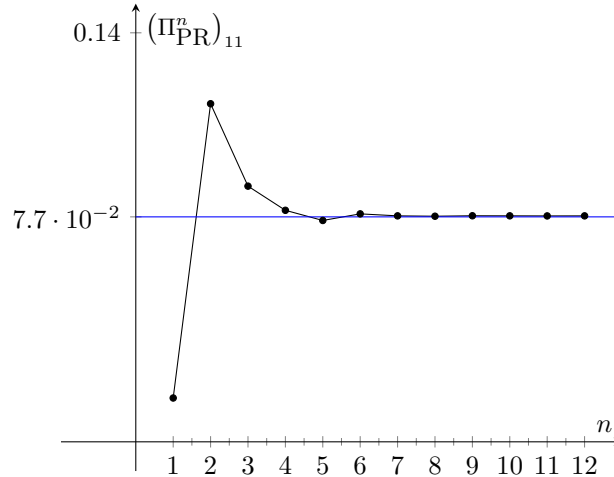
Ripetendo questo procedimento anche per le altre pagine, otteniamo un sistema lineare di 10 equazioni in incognite $\text{PR}(1), \dots, \text{PR}(10)$ equivalente a $\boldsymbol{\pi} = \boldsymbol{\pi}\Pi_{\text{PR}}$, ma tale sistema è stato ottenuto in un modo più semplice.

In questo esempio abbiamo determinato in modo esatto la distribuzione invariante $\boldsymbol{\pi}$ della catena di Markov X , ma in generale calcolare gli autovettori di una matrice $N \times N$ è molto costoso da un punto di vista computazionale e il metodo più efficiente per calcolare $\boldsymbol{\pi}$ è appunto quello di usare la convergenza menzionata nel Teorema 2.2.1, sfruttando la velocità di convergenza di tipo esponenziale. A titolo d'esempio questa è la matrice di transizione Π_{PR} , con $d = 0,85$, elevata a 12:

$$\Pi_{\text{PR}}^{12} \simeq \begin{pmatrix} 0,077 & 0,077 & 0,077 & 0,077 & 0,077 & 0,077 & 0,077 & 0,077 & 0,077 & 0,077 \\ 0,065 & 0,065 & 0,065 & 0,065 & 0,065 & 0,065 & 0,065 & 0,065 & 0,065 & 0,065 \\ 0,133 & 0,133 & 0,133 & 0,133 & 0,133 & 0,133 & 0,133 & 0,133 & 0,133 & 0,133 \\ 0,125 & 0,125 & 0,125 & 0,125 & 0,125 & 0,125 & 0,125 & 0,125 & 0,125 & 0,125 \\ 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 & 0,13 \\ 0,115 & 0,115 & 0,115 & 0,115 & 0,115 & 0,115 & 0,115 & 0,115 & 0,115 & 0,115 \\ 0,106 & 0,106 & 0,106 & 0,106 & 0,106 & 0,106 & 0,106 & 0,106 & 0,106 & 0,106 \\ 0,037 & 0,037 & 0,037 & 0,037 & 0,037 & 0,037 & 0,037 & 0,037 & 0,037 & 0,037 \\ 0,138 & 0,138 & 0,138 & 0,138 & 0,138 & 0,138 & 0,138 & 0,138 & 0,138 & 0,138 \\ 0,074 & 0,074 & 0,074 & 0,074 & 0,074 & 0,074 & 0,074 & 0,074 & 0,074 & 0,074 \end{pmatrix}.$$

Diamo ora una rappresentazione grafica della convergenza, da cui si intuisce che è di tipo esponenziale. Per fare ciò, a titolo d'esempio tracciamo il grafico dell'elemento nella

prima riga e prima colonna della matrice di transizione Π_{PR}^n , con $n = 1, \dots, 12$, ottenendo il seguente:



Come si vede dal grafico, la convergenza è molto veloce. Questo esempio può far congetturare che le costanti C e q che compaiono nella formula (2.2.2) del Teorema 2.2.1 non siano ottimali. Infatti, ripercorrendo la dimostrazione del Teorema 2.2.1, abbiamo:

$$q := (1 - \epsilon)^{\frac{1}{n_0}}, \quad C := (1 - \epsilon)^{-\frac{n_0 - 1}{n_0}},$$

dove $n_0 \in \mathbb{N}$ è tale che $\Pi_{\text{PR}}^{n_0}$ ha tutte le componenti strettamente positive (dalla definizione di catena di Markov regolare) e $\epsilon := \min_{i,j \in \mathcal{S}} \pi_{ij}^{(n_0)}$.

Nel nostro caso, con $d = 0,85$, abbiamo $\mathcal{S} = \{1, \dots, 10\}$, $n_0 = 1$ (poiché la matrice Π_{PR} ha tutte le componenti strettamente positive) e $\epsilon := \min_{i,j \in \mathcal{S}} \pi_{ij} = \frac{3}{200}$, da cui:

$$q = \frac{197}{200}, \quad C = 1.$$

Per tali valori di C e q , otteniamo:

$$|\pi_{ij}^{(12)} - \pi_j| \leq Cq^{12} = \left(\frac{197}{200}\right)^{12} \simeq 0,834, \quad \forall j \in \mathcal{S}.$$

Sappiamo invece che $|\pi_{ij}^{(12)} - \pi_j| \leq \frac{1}{1000}$, per ogni $j \in \mathcal{S}$, ma $Cq^n \leq \frac{1}{1000}$ solo se $n \geq 458$.

Bibliografia

- [1] Pierre Bremaud. *Markov Chains. Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer-Verlag New York, 1999. ISBN: 978-0-387-98509-1.
- [2] Sergey Brin e Lawrence Page. «The Anatomy of a Large-Scale Hypertextual Web Search Engine». In: *Computer Networks* 30 (1998), pp. 107–117. URL: <http://www-db.stanford.edu/~backrub/google.html>.
- [3] Nicola Fusco, Paolo Marcellini e Carlo Sbordone. *Lezioni di analisi matematica due*. Zanichelli, 2020. ISBN: 978-88-085-2020-3.
- [4] Andrea Pascucci. *Teoria della Probabilità. Variabili aleatorie e distribuzioni*. Springer-Verlag Mailand, 2020. ISBN: 978-88-470-3999-5.
- [5] Ernesto Salinelli e Franco Tomarelli. *Modelli Dinamici Discreti*. 3^a ed. Springer-Verlag Mailand, 2013. ISBN: 978-88-470-5503-2.
- [6] Albert N. Shiryaev. *Probability*. 2^a ed. Springer-Verlag New York, 1996.