Calcolo delle Probabilità e Statistica 2021/2022

Catene di Markov a tempo discreto

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

1 Processi stocastici

Iniziamo con l'introdurre una generalizzazione del concetto di variabile aleatoria, che si chiama *processo stocastico*[1] (o anche *processo casuale* o *processo aleatorio*).

Supponiamo di voler descrivere matematicamente una quantità numerica incerta il cui valore evolve nel tempo. Questo corrisponde ad una *famiglia di variabili aleatorie* indicizzate mediante un parametro che è appunto il *"tempo"*. Vediamo due esempi, a seconda che il tempo sia *"discreto"* o *"continuo"*.

Esempio 1.1. *Un'urna contiene 90 palline numerate da 1 a 90. Si estrae una pallina dall'urna, si registra il suo numero e la si reintroduce nell'urna. Quindi si itera la procedura. Poniamo*

$$X_n \;=\; \text{“numero della pallina estratta all'}n\text{-esima estrazione”}.$$

La quantità numerica (numero della pallina) che varia nel "tempo" (dato dall'ordine di estrazione) è rappresentata dalla famiglia (in tal caso, successione) di variabili aleatorie $(X_n)_n$*, indicizzate dal parametro* $n \in \mathbb{N}$.

Esempio 1.2. *Consideriamo un componente elettronico e sia*

$$X_t \;=\; \text{“temperatura del componente elettronico all'istante } t\text{”},$$

per ogni numero reale $t \geq 0$. *In tal caso la quantità numerica (temperatura) che varia nel tempo è rappresentata dalla famiglia di variabili aleatorie* $(X_t)_t$*, indicizzate dal parametro* $t \geq 0$.

Diamo quindi la definizione generale di processo stocastico, distinguendo a seconda della natura del parametro temporale.

Definizione 1.1. *Sia* $(\Omega, \mathbb{P})$ *uno spazio di probabilità.*

- *Si chiama* ***processo stocastico a tempo discreto*** *una successione di variabili aleatorie* $(X_n)_n$*, con* $n \in \mathbb{N}$*, tutte definite su* Ω.
- *Si chiama* ***processo stocastico a tempo continuo*** *una famiglia di variabili aleatorie* $(X_t)_t$*, con* $t \in [0, +\infty)$*, tutte definite su* Ω.

Osservazione. *Si noti che è possibile considerare processi stocastici su intervalli di tempo* finiti. *Ad esempio, un processo stocastico a tempo continuo sull'intervallo di tempo* $[0, 10]$ *è una famiglia di variabili aleatorie* $(X_t)_t$*, con* $t \in [0, 10]$.

Nel seguito considereremo solo processi stocastici a tempo discreto, in cui inoltre le variabili aleatorie $X_1, \ldots, X_n, \ldots$ saranno tutte *discrete* e, in particolare, con *supporto*

[1] Il termine *stocastico* deriva dal greco ed è sinonimo di *casuale* e *aleatorio*.

finito. Un esempio importante di processo stocastico a tempo discreto, che abbiamo già incontrato nello studio dei teoremi limite, è il seguente.

Esempio 1.3. *Sia $(X_n)_n$ una successione di variabili aleatorie indipendenti e identicamente distribuite (i.i.d.). Allora $(X_n)_n$ è un processo stocastico a tempo discreto.*

Il nostro obiettivo è studiare successioni $(X_n)_n$ più generali delle successioni i.i.d., in cui le due ipotesi ("indipendenti" e "identicamente distribuite") verranno indebolite. In particolare, per quanto riguarda la prima ipotesi ("indipendenti"), considereremo successioni $(X_n)_n$ in cui ci potrà essere una dipendenza tra le variabili aleatorie, anche se tale dipendenza dovrà avere una struttura particolare detta *"a catena"*: la variabile aleatoria X_{n+1} sarà influenzata direttamente solo da quella che la precede, ovvero X_n, e tramite essa sarà influenzata da tutte le variabili aleatorie precedenti. Più precisamente, una volta nota X_n, la conoscenza supplementare dei valori di $X_1, X_2, \ldots, X_{n-1}$ non darà alcuna ulteriore informazione riguardo il valore di X_{n+1}.

2 Catene di Markov a tempo discreto

I processi stocastici a tempo discreto che studieremo si chiamano *catene di Markov*[2] *a tempo discreto*, dove il termine "catene" fa proprio riferimento alla particolare struttura di dipendenza tra le variabili aleatorie del processo.

Definizione 2.1. *Si chiama* ***catena di Markov (a tempo discreto)***[a] *una successione di variabili aleatorie $(X_n)_n$ (ovvero, un processo stocastico a tempo discreto) che verifica le seguenti proprietà.*

1) Le variabili aleatorie $X_1, \ldots, X_n, \ldots$ sono discrete *e il loro supporto è contenuto nello* stesso *insieme $\mathcal{S}$ discreto (cioè finito o al più infinito numerabile), ovvero*

$$\mathcal{S}_{X_1} \subset \mathcal{S}, \qquad \cdots \qquad \mathcal{S}_{X_n} \subset \mathcal{S}, \qquad \cdots$$

$\mathcal{S}$ si chiama lo ***spazio degli stati*** *della catena di Markov.*

*2) (**Proprietà di Markov**: dipendenza a "catena") Per ogni scelta di $i_1, \ldots, i_{n-1}, i, j \in \mathcal{S}$ (non necessariamente distinti) vale l'uguaglianza*

$$\mathbb{P}(X_{n+1} = j | X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i) \;=\; \mathbb{P}(X_{n+1} = j | X_n = i).$$

[a] Dato che ci occuperemo solo di catene di Markov *a tempo discreto*, nel seguito parleremo solo di catene di Markov.

[2] Andrej Andreevic Markov (1856-1922) è stato un matematico russo, allievo di Chebyshev. Markov è noto soprattutto per essere stato uno dei primi matematici ad indagare a fondo i processi stocastici, introducendo in particolare i processi che oggi portano il suo nome.

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.

Osservazione 1. *Chiaramente l'uguaglianza*

$$\mathbb{P}(X_{n+1} = j | X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i) \ = \ \mathbb{P}(X_{n+1} = j | X_n = i)$$

deve valere solamente quando la probabilità condizionata è ben definita, ovvero solo se $\mathbb{P}(X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i) > 0$ *(da cui segue che* $\mathbb{P}(X_n = i) > 0$ *essendo* $\{X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i\} \subset \{X_n = i\}$*).*

Osservazione 2. *Per semplicità nel seguito utilizzeremo sempre e solo* $\mathcal{S}$ *invece dei supporti delle singole variabili aleatorie* $X_1, \ldots, X_n, \ldots$*, anche se ciascuna variabile aleatoria non assumerà necessariamente tutti i valori contenuti in* $\mathcal{S}$*.*

Osservazione 3. *La proprietà di Markov afferma che le variabili aleatorie* $X_1, \ldots, X_n, \ldots$ *non sono necessariamente indipendenti (come accade nel caso i.i.d.), tuttavia la struttura di dipendenza è abbastanza semplice dato che è appunto a "catena".*

Per comprendere meglio la proprietà di Markov, supponiamo di essere all'istante n, quindi di conoscere il valore che è stato assunto dalle variabili aleatorie $X_1, \ldots, X_{n-1}, X_n$. Indichiamo tali valori con $i_1, \ldots, i_{n-1}, i$, quindi

$$X_1 \ = \ i_1, \qquad X_2 \ = \ i_2, \qquad \cdots \qquad X_{n-1} \ = \ i_{n-1}, \qquad X_n \ = \ i.$$

I valori $i_1, \ldots, i_{n-1}, i$ rappresentano la *storia* del processo fino all'istante n; inoltre $i_1, \ldots, i_{n-1}$ sono i *valori passati*, mentre i è il *valore presente*; infine X_{n+1} è il *valore futuro*. Utilizzando questa terminologia, possiamo riscrivere la proprietà di Markov come segue:

$$\begin{aligned} &\mathbb{P}\big(\text{"il valore futuro è } j\text{"}\big|\text{"il valore presente è } i \text{ e i valori passati sono } i_1, \ldots, i_{n-1}\text{"}\big) \\ &\qquad\qquad = \ \mathbb{P}\big(\text{"il valore futuro è } j\text{"}\big|\text{"il valore presente è } i\text{"}\big) \end{aligned}$$

o, in modo ancora più sintetico,

$$\mathbb{P}(\text{"futuro"}|\text{"presente e passato"}) \ = \ \mathbb{P}(\text{"futuro"}|\text{"presente"}).$$

In altri termini, nelle catene di Markov la dipendenza dal passato dei valori futuri viene riassunta dal valore presente. Chiaramente è possibile considerare processi stocastici con strutture di dipendenza più complesse di quella delle catene di Markov, ma noi non ce ne occuperemo.

3 Catene di Markov omogenee e a stati finiti

Nel seguito ci occuperemo solamente di una classe particolare e molto importante di catene di Markov, ovvero le *catene di Markov omogenee e a stati finiti*, che ora definiamo.

Definizione 3.1. *Sia $(X_n)_n$ una catena di Markov.*

- *Si dice che $(X_n)_n$ è* ***omogenea (nel tempo)*** *se la probabilità di transizione non dipende da n. In tal caso, si scrive π_{ij} invece di $\pi_{ij}(n)$ e si dice che π_{ij} è la* ***probabilità di transizione dallo stato i allo stato j.***
- *Si dice che $(X_n)_n$ è* ***a stati finiti*** *se lo spazio di stato $\mathcal{S}$ è un insieme finito. In tal caso indicheremo con N la cardinalità di $\mathcal{S}$. Spesso supporremo che $\mathcal{S}$ sia dato da $\mathcal{S} = \{1, 2, \ldots, N\}$ oppure $\mathcal{S} = \{0, 1, \ldots, N-1\}$.*

NOTAZIONE. *Nel seguito, anche se non esplicitamente indicato, quando scriveremo* catena di Markov *intenderemo sempre* catena di Markov omogenea e a stati finiti.

La struttura di dipendenza a catena di $(X_n)_n$, nel caso omogeneo e a stati finiti, è completamente descritta da una matrice quadrata di ordine N, detta *matrice di transizione.*

Definizione 3.2. *Sia $(X_n)_n$ una catena di Markov omogenea e a stati finiti. Si chiama* ***matrice di transizione*** *la matrice $N \times N$ (dove N è la cardinalità di $\mathcal{S}$), indicata con Π, le cui componenti sono le probabilità di transizione:*

$$\pi_{ij} \;=\; \mathbb{P}(X_{n+1} = j | X_n = i), \qquad \text{per ogni } i, j \in \mathcal{S},$$

con n qualunque, dato che la catena è omogenea. In altri termini, la probabilità di transizione π_{ij} corrisponde all'elemento nella riga i e colonna j della matrice Π.

OSSERVAZIONE. *Se lo spazio degli stati $\mathcal{S}$ non è l'insieme $\{1, \ldots, N\}$, per scrivere una matrice di transizione Π bisogna prima fissare un ordinamento degli stati in $\mathcal{S}$ (ossia decidere quale stato corrisponde alla prima riga, quale alla seconda, e così via).*

Si noti che ogni riga di Π corrisponde alle probabilità (si pensi ad esempio alla riga i):

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

La riga i di Π corrisponde quindi alla "densità discreta di X_{n+1} sapendo che $X_n = i$". Questo implica che ogni elemento di Π deve essere un numero appartenente all'intervallo $[0, 1]$, in quanto corrisponde ad una probabilità (condizionata). Inoltre, la somma degli elementi di una qualsiasi riga deve essere uguale a 1. In altri termini, Π deve verificare le proprietà riportate nel seguente teorema.

Teorema 3.1. *Sia Π una matrice di transizione di una catena di Markov $(X_n)_n$. Allora Π è tale che:*

1) $0 \leq \pi_{ij} \leq 1$, *per ogni i, j;*

2) *la somma degli elementi di ogni riga vale 1, ovvero*

$$\sum_{j=1}^{N} \pi_{ij} \;=\; 1, \qquad \text{per ogni riga } i.$$

Dimostrazione. La proprietà 1) segue direttamente dalla definizione di π_{ij}, in quanto ciascun π_{ij} è una probabilità (condizionata). Per quanto riguarda 2), essa è una conseguenza della *formula delle probabilità totali*. Infatti, sia i una riga qualsiasi. Poniamo (scegliendo un istante n qualsiasi, dato che la catena di Markov è omogenea)

$$A = \{X_n = i\}, \qquad B_j = \{X_{n+1} = j\}, \text{ per ogni } j \in \mathcal{S} = \{1, \dots, N\}.$$

Allora gli eventi $B_1, \dots, B_N$ costituiscono una partizione di Ω. Quindi, per la formula delle probabilità totali, si ha

$$\mathbb{P}(A) = \sum_{j=1}^{N} \mathbb{P}(A \cap B_j),$$

che possiamo riscrivere come segue

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

Utilizziamo quest'ultima uguaglianza, indicata con $(*)$, per dimostrare la proprietà 2). Abbiamo che

$$\begin{aligned}
\sum_{j=1}^{N} \pi_{ij} = \sum_{j=1}^{N} \mathbb{P}(X_{n+1} = j | X_n = i) &= \sum_{j=1}^{N} \frac{\mathbb{P}(X_{n+1} = j, X_n = i)}{\mathbb{P}(X_n = i)} \\
&= \frac{1}{\mathbb{P}(X_n = i)} \sum_{j=1}^{N} \mathbb{P}(X_{n+1} = j, X_n = i) \\
&\underset{\substack{\uparrow \\ (*)}}{=} \frac{\mathbb{P}(X_n = i)}{\mathbb{P}(X_n = i)} = 1.
\end{aligned}$$

□

Rappresentazione grafica di una catena di Markov. Una catena di Markov può essere rappresentata graficamente tramite un *grafo orientato* costruito nel modo seguente:

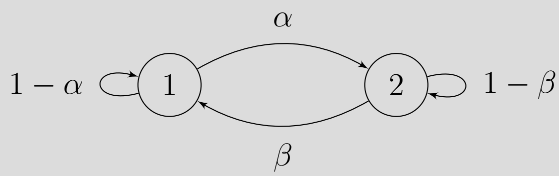
- ogni stato $i \in \mathcal{S}$ corrisponde ad un nodo del grafo;
- ogni probabilità di transizione π_{ij}, se strettamente positiva, corrisponde ad un arco orientato (una freccia) dal nodo i al nodo j (non si disegnano invece le frecce corrispondenti a probabilità di transizione nulle);
- si riporta il valore di π_{ij} sull'arco corrispondente.

In tal caso si dice che la successione di variabili aleatorie $(X_n)_n$ è una *passeggiata aleatoria* (in inglese *random walk*) sul grafo. Vediamo due esempi.

Esempio 3.1. *Si consideri una catena di Markov* $(X_n)_n$ *con spazio degli stati* $\mathcal{S} = \{1, 2\}$ *e matrice di transizione*

$$\Pi \;=\; \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

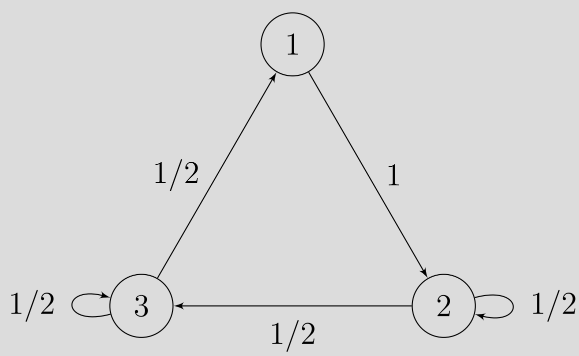
dove $0 \le \alpha \le 1$ *e* $0 \le \beta \le 1$ *sono due parametri fissati. Questa è la matrice di transizione più generale possibile per una catena di Markov a due stati. Il grafo ad essa associato è il seguente:*



Esempio 3.2. *Si consideri una catena di Markov* $(X_n)_n$ *con spazio degli stati* $\mathcal{S} = \{1, 2, 3\}$ *e matrice di transizione*

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

Il grafo ad essa associato è il seguente:



3.1 Probabilità di transizione in più passi

Sia $(X_n)_n$ una catena di Markov (omogenea e a stati finiti) con matrice di transizione Π. Dalla definizione di $(X_n)_n$, sappiamo quanto vale la probabilità condizionata

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

infatti è data da π_{ij}. Quanto vale invece la probabilità condizionata

$$\mathbb{P}(X_{n+2} = j | X_n = i)\,?$$

Più in generale, quanto vale

$$\mathbb{P}(X_{n+m} = j | X_n = i)\,?$$

Tale probabilità (che nel caso omogeneo non dipende da n) si indica con $\pi_{ij}^{(m)}$ e si chiama **probabilità di transizione dallo stato i allo stato j in m passi**.

Definizione 3.3. *Sia $(X_n)_n$ una catena di Markov omogenea e a stati finiti. Per ogni intero $m \geq 0$, poniamo (con n qualsiasi)*

$$\pi_{ij}^{(m)} = \mathbb{P}(X_{n+m} = j | X_n = i), \qquad \text{per ogni } i, j \in \mathcal{S}.$$

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

Quando $m = 0$ oppure $m = 1$, la probabilità di transizione ha un'espressione particolare, come descritto nella seguente proposizione.

Proposizione 3.1. *Sia $(X_n)_n$ una catena di Markov omogenea e a stati finiti.*

- *Per $m = 0$, $\pi_{ij}^{(0)}$ è data da:*

$$\pi_{ij}^{(0)} = \begin{cases} 1, & \text{se } i = j, \\ 0, & \text{se } i \neq j. \end{cases}$$

  Quindi la matrice di componenti $\pi_{ij}^{(0)}$ corrisponde alla matrice identità $N \times N$, indicata con I_N.

- *Per $m = 1$, $\pi_{ij}^{(1)}$ è data da:*

$$\pi_{ij}^{(1)} = \pi_{ij}.$$

  Quindi la matrice di componenti $\pi_{ij}^{(1)}$ corrisponde alla matrice di transizione Π.

Dimostrazione.

- Sia $m = 0$, allora

$$\pi_{ij}^{(0)} = \mathbb{P}(X_n = j | X_n = i),$$

  dove n è un istante qualsiasi, dato che la catena di Markov è omogenea. Per definizione di probabilità condizionata, abbiamo che

$$\mathbb{P}(X_n = j | X_n = i) = \frac{\mathbb{P}(\{X_n = j\} \cap \{X_n = i\})}{\mathbb{P}(X_n = i)}.$$

  Si noti 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

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

- Sia $m = 1$. Segue direttamente dalle definizioni di π_{ij} e $\pi_{ij}^{(1)}$ che tali probabilità (condizionali) coincidono.

$\square$

Il seguente teorema fornisce una formula per il calcolo di $\pi_{ij}^{(m)}$.

Teorema 3.2. *Sia $(X_n)_n$ una catena di Markov omogenea e a stati finiti. Per ogni intero m, la matrice di componenti $\pi_{ij}^{(m)}$ è data da*

$$\underbrace{\Pi \cdots \Pi}_{m \text{ volte}} \;=\; \Pi^m.$$

Osservazione. *Si noti che, in particolare, il Teorema 3.2 vale anche per $m = 1$, dato che $\Pi^1 = \Pi$, e per $m = 0$, dato che $\Pi^0 = I_N$.*

Dimostrazione del Teorema 3.2 nel caso $m = 2$. Dobbiamo dimostrare che $\pi_{ij}^{(2)}$ coincide con l'elemento nella i-esima riga e j-esima colonna della matrice Π^2, quindi

$$\pi_{ij}^{(2)} \;=\; \sum_{k=1}^{N} \pi_{ik}\pi_{kj}.$$

Per mostrare la validità di questa formula, consideriamo tutti i cammini che portano da i a j in due passi. Ci sono esattamente N cammini di questo tipo, che sono dati da:

$$i \to 1 \to j; \qquad i \to 2 \to j; \qquad \cdots \qquad i \to N \to j.$$

Consideriamo uno qualunque di questi cammini, ad esempio quello passante per lo stato k, ovvero $i \to k \to j$. La probabilità di percorrerlo è pari al prodotto delle probabilità $\pi_{ik}\pi_{kj}$. Infatti

$$\begin{aligned}
&\mathbb{P}(\text{“passare per } k \text{ e poi per } j\text{”}|\text{“partire da } i\text{”}) \underset{\substack{\uparrow \\ n \text{ qualunque, per omog.}}}{=} \mathbb{P}(\{X_{n+2} = j\} \cap \{X_{n+1} = k\} | X_n = i) \\
&= \frac{\mathbb{P}(\{X_{n+2} = j\} \cap \{X_{n+1} = k\} \cap \{X_n = i\})}{\mathbb{P}(X_n = i)} \\
&= \frac{\mathbb{P}(\{X_{n+1} = k\} \cap \{X_n = i\})}{\mathbb{P}(X_n = i)} \frac{\mathbb{P}(\{X_{n+2} = j\} \cap \{X_{n+1} = k\} \cap \{X_n = i\})}{\mathbb{P}(\{X_{n+1} = k\} \cap \{X_n = i\})} \\
&= \mathbb{P}(X_{n+1} = k | X_n = i)\, \mathbb{P}(X_{n+2} = j | \{X_{n+1} = k\} \cap \{X_n = i\}) \\
&\underset{\substack{\uparrow \\ \text{propr. Markov}}}{=} \mathbb{P}(X_{n+1} = k | X_n = i)\, \mathbb{P}(X_{n+2} = j | X_{n+1} = k) \;=\; \pi_{ik}\pi_{kj}.
\end{aligned}$$

Infine, $\pi_{ij}^{(2)}$ è dato dalla somma delle probabilità dei cammini che portano da i a j in due passi. Quindi

$$\pi_{ij}^{(2)} \;=\; \sum_{k=1}^{N} \pi_{ik}\pi_{kj}.$$

$\square$

Calcolo diretto di $\pi_{ij}^{(m)}$**.** Dal Teorema 3.2 sappiamo che per calcolare $\pi_{ij}^{(m)}$ dobbiamo prima trovare la matrice Π^m, dopodiché $\pi_{ij}^{(m)}$ è l'elemento nella riga i e colonna j di questa matrice. Vediamo ora invece un modo alternativo per il calcolo di $\pi_{ij}^{(m)}$, più diretto, che si basa sull'utilizzo del grafo orientato associato a $(X_n)_n$. Vediamolo nel caso $m = 2$. Sappiamo che $\pi_{ij}^{(2)}$ è dato dalla formula

$$\pi_{ij}^{(2)} = \sum_{k=1}^{N} \pi_{ik}\pi_{kj}.$$

Il prodotto $\pi_{ik}\pi_{kj}$ corrisponde alla probabilità di andare da i a j in due passi passando per il nodo k, cioè di percorrere il cammino che indichiamo brevemente con $i \to k \to j$. Generalmente, solo alcuni di questi cammini hanno probabilità positiva (cioè solo per alcuni si ha che $\pi_{ik}\pi_{kj} > 0$). Per individuarli, il modo più facile è utilizzare il grafo associato alla catena di Markov, sfruttando il fatto che sul grafo non sono riportate le frecce corrispondenti a probabilità nulle. Se ad esempio vediamo dal grafo che esistono solo due cammini possibili, dati da $i \to k_1 \to j$ e $i \to k_2 \to j$, allora

$$\pi_{ij}^{(2)} = \pi_{ik_1}\pi_{k_1 j} + \pi_{ik_2}\pi_{k_2 j}.$$

In generale, per calcolare $\pi_{ij}^{(m)}$, con m qualunque, si procede come segue:

- a partire dal grafo, si trovano tutti i cammini che portano da i a j in m passi;
- la probabilità di ogni cammino è il prodotto delle probabilità lungo gli archi del cammino stesso;
- $\pi_{ij}^{(m)}$ è la somma delle probabilità dei cammini che portano da i a j in m passi.

3.2 Classi comunicanti

Data una catena di Markov $(X_n)_n$ (omogenea e a stati finiti), è possibile classificare gli stati in modo tale da partizionare lo spazio degli stati $\mathcal{S}$ in sottoinsiemi chiamati *classi comunicanti*.

Definizione 3.4. *Sia $(X_n)_n$ una catena di Markov omogenea e a stati finiti. Fissiamo due stati $i, j \in \mathcal{S}$ (non necessariamente $i \neq j$).*
Si dice[a] *che j* ***è accessibile da*** *i se esiste $m \geq 0$ tale che*

$$\pi_{ij}^{(m)} > 0.$$

In tal caso scriviamo

$$i \rightsquigarrow j.$$

[a] *In termini di grafo orientato significa che il nodo i è* ***connesso*** *con il nodo j.*

Osservazione. *Dato che, per definizione, $\pi_{ii}^{(0)} = 1$, è sempre vero che i è accessibile da i stesso. Infatti è banalmente vero che se parto dalla stato i allora accedo allo stato i in $m = 0$ passi. Quindi vale sempre che $i \rightsquigarrow i$.*

Nel caso $i \neq j$, come chiarito nel teorema che segue, si ha che $i \rightsquigarrow j$ se e solo se esiste un cammino *di probabilità positiva* (che dunque può essere determinato a partire dal grafo) che conduce da i a j.

Teorema 3.3. *Le due affermazioni seguenti sono equivalenti se $i \neq j$:*

a) $i \rightsquigarrow j$*;*

b) *esiste un intero $m \geq 1$ ed esiste un cammino $i_1 \to i_2 \to i_3 \to \cdots \to i_{m+1}$ in m passi tale che $i_1 = i$, $i_{m+1} = j$ e*

$$\pi_{i_1 i_2} \pi_{i_2 i_3} \cdots \pi_{i_m i_{m+1}} \; > \; 0.$$

Dimostrazione del Teorema 3.3 nel caso $m = 2$. Dobbiamo dimostrare che le seguenti affermazioni sono equivalenti se $i \neq j$:

a) $i \rightsquigarrow j$, ovvero $\pi_{ij}^{(2)} > 0$;

b) esiste un cammino $i \to k \to j$, che porta da i a j in due passi, tale che

$$\pi_{ik}\pi_{kj} \; > \; 0.$$

Sappiamo dal Teorema 3.2 che

$$\pi_{ij}^{(2)} \; = \; \sum_{k=1}^{N} \pi_{ik}\pi_{kj}$$

Concludiamo dunque che $\pi_{ij}^{(2)} > 0$ se e solo se almeno un addendo della sommatoria è positivo, ovvero se e solo se esiste un k tale che $\pi_{ik}\pi_{kj} > 0$. □

Definizione 3.5. *Sia $(X_n)_n$ una catena di Markov omogenea e a stati finiti. Fissiamo due stati $i, j \in \mathcal{S}$ (non necessariamente $i \neq j$).*

Gli stati i e j si dicono[a] ***comunicanti*** *se $i \rightsquigarrow j$ e $j \rightsquigarrow i$. In tal caso scriviamo*

$$i \leftrightsquigarrow j.$$

Si chiama[b] ***classe comunicante*** *un sottoinsieme di $\mathcal{S}$ costituito da tutti gli stati tra loro comunicanti.*

[a] *In termini di grafo orientato significa che il nodo i è* ***fortemente connesso*** *con il nodo j.*

[b] *In termini di grafo orientato la classe comunicante prende il nome di* ***componente fortemente connessa****.*

Osservazione 1. *Dall'osservazione precedente sappiamo che $i \rightsquigarrow i$ per qualunque stato i. Dunque è chiaro che* ***ogni stato $i \in \mathcal{S}$ è comunicante con se stesso****, ovvero $i \leftrightsquigarrow i$. Ciò implica che ogni stato $i \in \mathcal{S}$ appartiene ad una classe comunicante (tale classe potrebbe eventualmente contenere solo lo stato i stesso).*

Osservazione 2. *Dal Teorema 3.3 segue che le due affermazioni seguenti sono equivalenti se $i \neq j$:*

a) $i \leftrightsquigarrow j$;

b) *esiste un* cammino chiuso *che passa per i e j (per* cammino chiuso *si intende*[3] *un cammino in cui stato di partenza e stato di arrivo coincidono).*

Osservazione 3. *La relazione "comunicante con" è una* ***relazione di equivalenza*** *sullo spazio degli stati $\mathcal{S}$, ovvero verifica le seguenti proprietà:*

- ***riflessività****: $i \leftrightsquigarrow i$, per ogni stato $i \in \mathcal{S}$;*
- ***simmetria****: se $i \leftrightsquigarrow j$ allora $j \leftrightsquigarrow i$;*
- ***transitività****: se $i \leftrightsquigarrow j$ e $j \leftrightsquigarrow k$ allora $i \leftrightsquigarrow k$.*

*Infatti, è riflessiva per l'*Osservazione 1 *riportata qui sopra. Inoltre, utilizzando quanto affermato nell'*Osservazione 2*, è facile verificare che è anche simmetrica e transitiva. Dunque l'insieme $\mathcal{S}$ può essere partizionato nelle corrispondenti* classi di equivalenza*, che sono appunto le* ***classi comunicanti****. Da ciò segue che* ***ogni stato $i \in \mathcal{S}$ appartiene ad una e una sola classe comunicante*** *(non è in particolare possibile che uno stesso stato appartenga contemporaneamente a due classi distinte).*

Definizione 3.6. *Sia $(X_n)_n$ una catena di Markov omogenea e a stati finiti.*
Si dice che $(X_n)_n$ è ***irriducibile*** *se esiste un'unica classe comunicante, che è quindi data dall'insieme $\mathcal{S}$ stesso.*

3.3 Legge di X_n

Sia $(X_n)_n$ una catena di Markov (omogenea e a stati finiti) con matrice di transizione Π. Ci poniamo ora il seguente problema.

Qual è la legge della variabile aleatoria X_n?

Sappiamo che X_n è una variabile aleatoria *discreta*, quindi è sufficiente determinare supporto e densità discreta di X_n. Come abbiamo detto in precedenza, come supporto prendiamo $\mathcal{S}$ (anche se, in generale, $\mathcal{S}$ contiene più valori di quelli che effettivamente X_n assume). Resta da determinare la densità discreta p_{X_n}. Supponiamo che lo spazio degli stati $\mathcal{S}$ sia dato dall'insieme $\{1, 2, \ldots, N\}$. Allora determinare la densità discreta p_{X_n} significa conoscere la tabella

X_n	1	2	$\cdots$	N
p_{X_n}	$p_{X_n}(1)$	$p_{X_n}(2)$	$\cdots$	$p_{X_n}(N)$

Più precisamente, determinare la densità discreta p_{X_n} significa conoscere il vettore riga $\vec{\boldsymbol{p}}_{\boldsymbol{X_n}}$ dato da

$$\vec{\boldsymbol{p}}_{\boldsymbol{X_n}} = \begin{bmatrix} p_{X_n}(1) & p_{X_n}(2) & \cdots & p_{X_n}(N) \end{bmatrix} = \begin{bmatrix} \mathbb{P}(X_n = 1) & \mathbb{P}(X_n = 2) & \cdots & \mathbb{P}(X_n = N) \end{bmatrix}.$$

[3] Ricordiamo che per *ciclo* si intende invece un particolare tipo di cammino chiuso, in cui gli stati intermedi (ovvero tutti gli stati tranne quelli di partenza e di arrivo, che coincidono) sono *distinti* tra loro.

Notazione. *Nel seguito, anche se non esplicitamente detto, quando parleremo di distribuzione (o legge) di X_n ci riferiremo sempre al vettore riga $\vec{\boldsymbol{p}}_{\boldsymbol{X_n}}$, che come abbiamo visto descrive completamente la densità discreta di X_n, e quindi anche la sua distribuzione.*

Torniamo al problema che ci siamo posti inizialmente, ovvero determinare la distribuzione di X_n (quindi $\vec{\boldsymbol{p}}_{\boldsymbol{X_n}}$). Per trovare tale distribuzione non è sufficiente conoscere la matrice di transizione Π, dobbiamo anche sapere qual è la *distribuzione iniziale* della catena di Markov, cioè la distribuzione di X_1 (quindi $\vec{\boldsymbol{p}}_{\boldsymbol{X_1}}$).

Teorema 3.4. *Sia $(X_n)_n$ una catena di Markov omogenea e a stati finiti. Allora la distribuzione di X_n è data dalla seguente formula:*

$$\vec{\boldsymbol{p}}_{\boldsymbol{X_n}} \;=\; \vec{\boldsymbol{p}}_{\boldsymbol{X_1}}\,\Pi^{n-1}, \qquad \textit{per ogni } n = 1, 2, \ldots$$

Si noti in particolare che

$$p_{X_n}(j) \;=\; \mathbb{P}(X_n = j) \;=\; \sum_{i=1}^{N} p_{X_1}(i)\,\pi_{ij}^{(n-1)},$$

per ogni $j \in \mathcal{S}$.

Dimostrazione. Dobbiamo dimostrare che, per ogni $j \in \mathcal{S}$, vale la seguente formula:

$$\mathbb{P}(X_n = j) \;=\; \sum_{i=1}^{N} p_{X_1}(i)\,\pi_{ij}^{(n-1)}.$$

Tale formula è vera in quanto è una diretta conseguenza della *formula delle probabilità totali*. Infatti

$$\mathbb{P}(X_n = j) \;=\; \sum_{i=1}^{N} \underbrace{\mathbb{P}(X_n = j | X_1 = i)}_{\pi_{ij}^{(n-1)}}\,\underbrace{\mathbb{P}(X_1 = i)}_{p_{X_1}(i)}.$$

□

3.4 Distribuzione invariante

Sia X una variabile aleatoria discreta con supporto (contenuto in) $\mathcal{S}$. Sappiamo che la distribuzione (o legge) di X è completamente descritta dal vettore riga $\vec{\boldsymbol{p}}_{\boldsymbol{X}}$, che contiene i valori assunti dalla densità discreta di X. Per semplificare la notazione, indichiamo questo vettore con $\vec{\boldsymbol{\pi}}$. Si noti che il vettore $\vec{\boldsymbol{\pi}}$ verifica le seguenti proprietà:

1) ogni sua componente è compresa tra 0 e 1;

2) la somma delle sue componenti è uguale a 1.

Definizione 3.7. *Siano Π una matrice di transizione e $\vec{\boldsymbol{\pi}}$ un vettore che verifica le proprietà 1) e 2) qui sopra riportate.*
Si dice che $\vec{\boldsymbol{\pi}}$ è una ***distribuzione invariante*** *o* ***stazionaria*** *o* ***di equilibrio*** *(per Π) se*

$$\vec{\boldsymbol{\pi}} \;=\; \vec{\boldsymbol{\pi}}\,\Pi.$$

Osservazione. *Si noti che* $\vec{\boldsymbol{\pi}}$ *è una distribuzione invariante per* Π *se e solo se* $\vec{\boldsymbol{\pi}}$ *è un autovettore (che verifica le proprietà 1) e 2)) per la* ***matrice trasposta*** *di* Π *con autovalore* 1.

Il termine "invariante" deriva dal seguente teorema, in cui si dimostra che se la distribuzione iniziale della catena di Markov è invariante allora le variabili aleatorie $X_1, \dots, X_n, \dots$ sono *identicamente distribuite.*

Teorema 3.5. *Sia* $(X_n)_n$ *una catena di Markov omogenea e a stati finiti con matrice di transizione* Π.
Supponiamo che la distribuzione di X_1 *sia invariante, ovvero*

$$\vec{\boldsymbol{p}}_{\boldsymbol{X_1}} \;=\; \vec{\boldsymbol{p}}_{\boldsymbol{X_1}}\,\Pi.$$

Allora la distribuzione di X_n *(qualunque sia* n*) è ancora data da* $\vec{\boldsymbol{p}}_{\boldsymbol{X_1}}$.

Dimostrazione. Dal Teorema 3.4, sappiamo che la distribuzione (intesa come densità discreta) di X_n è data da

$$\vec{\boldsymbol{p}}_{\boldsymbol{X_n}} \;=\; \vec{\boldsymbol{p}}_{\boldsymbol{X_1}}\,\Pi^{n-1} \;=\; \underbrace{\vec{\boldsymbol{p}}_{\boldsymbol{X_1}}\,\Pi}_{=\vec{\boldsymbol{p}}_{\boldsymbol{X_1}}}\,\Pi^{n-2} \;=\; \vec{\boldsymbol{p}}_{\boldsymbol{X_1}}\,\Pi^{n-2} \;=\; \cdots \;=\; \vec{\boldsymbol{p}}_{\boldsymbol{X_1}}.$$

□

Come vedremo nella prossima sezione, il concetto di distribuzione invariante di una catena di Markov gioca un ruolo importante in quanto è legato al comportamento della catena stessa per tempi lunghi.

4 Algoritmo PageRank

In questa sezione finale vogliamo descrivere come funziona il motore di ricerca Google (nella sua prima versione del 1997), o meglio come funziona l'algoritmo PageRank. Prima però abbiamo bisogno di introdurre ancora alcuni strumenti della teoria delle catene di Markov.

4.1 Interpretazione della distribuzione invariante

Iniziamo col dare la seguente definizione.

Definizione 4.1. *Sia* $(X_n)_n$ *una catena di Markov omogenea e a stati finiti con matrice di transizione* Π.
Si dice che $(X_n)_n$ *è* ***regolare*** *se esiste* n_0 *tale che*

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

ovvero se la matrice

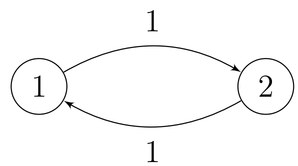
$$\underbrace{\Pi \cdots \Pi}_{n_0 \ volte} \;=\; \Pi^{n_0}$$

ha tutte le componenti strettamente positive.

Osservazione. *Si noti che se una catena di Markov è regolare allora si può andare da un qualunque stato i ad un qualunque altro stato j in al più n_0 passi (dato che $\pi_{ij}^{(n_0)} > 0$ per ogni $i, j \in \mathcal{S}$). Dunque, in tal caso, $(X_n)_n$ ha un'unica classe comunicante (che è quindi $\mathcal{S}$). In altri termini, se una catena di Markov è regolare allora è irriducibile. Non è vero il viceversa, come dimostra il seguente esempio. Sia $(X_n)_n$ la catena di Markov con matrice di transizione*

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

e, quindi, grafo orientato



Allora $(X_n)_n$ è irriducibile ma non è regolare. Infatti vale che $\Pi^2 = I_2$, dove I_2 è la matrice identità 2×2. Quindi Π^n è uguale a Π se n è dispari ed è uguale a I_2 se n è pari. Perciò Π^n non ha mai tutte le componenti strettamente positive, cioè $(X_n)_n$ non è regolare.

Se una catena di Markov è regolare vale il seguente risultato fondamentale.

Teorema 4.1 (di convergenza all'equilibrio o ergodico)**.** *Sia $(X_n)_n$ una catena di Markov omogenea e a stati finiti con matrice di transizione Π.*
Se $(X_n)_n$ è regolare allora esiste un'unica distribuzione invariante $\vec{\boldsymbol{\pi}}$ tale che, qualunque sia $i \in \mathcal{S}$,

$$\lim_{n \to +\infty} \pi_{ij}^{(n)} \;=\; \pi_j, \qquad \forall\, j \in \mathcal{S}. \tag{4.1}$$

Inoltre, la velocità di convergenza è esponenziale:

$$|\pi_{ij}^{(n)} - \pi_j| \;\le\; C\, q^n, \tag{4.2}$$

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

Dimostrazione. Per ogni $j \in \mathcal{S}$, poniamo

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

Per il Teorema 3.2, abbiamo che

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

da cui ricaviamo la seguente disuguaglianza:

$$m_j^{(n+1)} \;=\; \min_{i \in \mathcal{S}} \pi_{ij}^{(n+1)} \;=\; \min_{i \in \mathcal{S}} \left(\sum_{k \in \mathcal{S}} \pi_{ik} \pi_{kj}^{(n)} \right) \;\ge\; \min_{i \in \mathcal{S}} \left(\sum_{k \in \mathcal{S}} \pi_{ik} \min_{h \in \mathcal{S}} \pi_{hj}^{(n)} \right). \tag{4.3}$$

Dalle proprietà della matrice di transizione, sappiamo che $\sum_{k\in\mathcal{S}} \pi_{ik} = 1$, da cui anche $\min_{i\in\mathcal{S}} \sum_{k\in\mathcal{S}} \pi_{ik} = 1$. Pertanto, usando la definizione di $m_j^{(n)}$, da (4.3) otteniamo

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

Quindi la successione $(m_j^{(n)})_{n\in\mathbb{N}}$ è monotona crescente, per ogni $j \in \mathcal{S}$. Con un ragionamento analogo si dimostra che la successione

$$(M_j^{(n)})_{n\in\mathbb{N}}$$

è monotona decrescente. Perciò, per provare (4.1) è sufficiente mostrare che

$$M_j^{(n)} - m_j^{(n)} \overset{n\to+\infty}{\longrightarrow} 0, \qquad \forall\, j \in \mathcal{S}.$$

Poiché la catena di Markov è regolare, esiste $n_0 \in \mathbb{N}$ tale che: $\pi_{ij}^{(n_0)} > 0$, $\forall\, i, j \in \mathcal{S}$. Poniamo $\varepsilon := \min_{i,j} \pi_{ij}^{(n_0)}$ (ovviamente si ha $\varepsilon > 0$). Allora, per il Teorema 3.2 abbiamo che

$$\begin{aligned}
\pi_{ij}^{(n_0+n)} \;&=\; \sum_{k\in\mathcal{S}} \pi_{ik}^{(n_0)}\pi_{kj}^{(n)} \;=\; \sum_{k\in\mathcal{S}} \big(\pi_{ik}^{(n_0)} - \varepsilon\pi_{jk}^{(n)}\big)\pi_{kj}^{(n)} + \varepsilon \sum_{k\in\mathcal{S}} \pi_{jk}^{(n)}\pi_{kj}^{(n)} \\
&=\; \sum_{k\in\mathcal{S}} \big(\pi_{ik}^{(n_0)} - \varepsilon\pi_{jk}^{(n)}\big)\pi_{kj}^{(n)} + \varepsilon\pi_{jj}^{(2n)}.
\end{aligned} \tag{4.4}$$

Ma, poiché $\pi_{jk}^{(n)} \in [0,1]$, dalla definizione di ε si ha: $\pi_{ik}^{(n_0)} - \varepsilon\pi_{jk}^{(n)} \geq 0$. Pertanto, da (4.4), usando la definizione di $m_j^{(n)}$ e il Teorema 3.1, otteniamo

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

Per l'arbitrarietà di i si ottiene

$$m_j^{(n_0+n)} \;\geq\; m_j^{(n)}(1-\varepsilon) + \varepsilon\pi_{jj}^{(2n)}. \tag{4.5}$$

Analogamente, da (4.4), usando la definizione di $M_j^{(n)}$ e il Teorema 3.1, si ottiene

$$M_j^{(n_0+n)} \;\leq\; M_j^{(n)}(1-\varepsilon) + \varepsilon\pi_{jj}^{(2n)}. \tag{4.6}$$

Combinando (4.5) e (4.6), otteniamo

$$M_j^{(n_0+n)} - m_j^{(n_0+n)} \;\leq\; \big(M_j^{(n)} - m_j^{(n)}\big)(1-\varepsilon)$$

e, di conseguenza,

$$\begin{aligned}
M_j^{(hn_0+n)} - m_j^{(hn_0+n)} \;&\leq\; \big(M_j^{((h-1)n_0+n)} - m_j^{((h-1)n_0+n)}\big)(1-\varepsilon) \\
&\leq\; \big(M_j^{((h-2)n_0+n)} - m_j^{((h-2)n_0+n)}\big)(1-\varepsilon)(1-\varepsilon) \\
&\leq\; \cdots \;\leq\; \big(M_j^{(n_0+n)} - m_j^{(n_0+n)}\big)(1-\varepsilon)^{h-1} \;\leq\; \big(M_j^{(n)} - m_j^{(n)}\big)(1-\varepsilon)^h.
\end{aligned} \tag{4.7}$$

Poiché $0 < \varepsilon \leq 1$, si ha che $(1-\varepsilon)^h \to 0$, quando $h \to +\infty$, quindi

$$M_j^{(hn_0+n)} - m_j^{(hn_0+n)} \overset{h\to+\infty}{\longrightarrow} 0.$$

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

$$\lim_{n\to+\infty} M_j^{(n)} - m_j^{(n)} \overset{n\to+\infty}{\longrightarrow} 0, \qquad \forall\, j \in \mathcal{S}.$$

Poniamo

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

Questo conclude la dimostrazione di (4.1).
Dimostriamo infine la convergenza esponenziale (4.2). Poiché $(M_j^{(n)} - m_j^{(n)})_{n\in\mathbb{N}}$ è monotona decrescente, dalla (4.7) e ponendo $h = [n/n_0]$ (ossia, h è la parte intera di n/n_0; inoltre, si noti che $hn_0 \leq n$), abbiamo

$$M_j^{(n)} - m_j^{(n)} \;\leq\; M_j^{(hn_0)} - m_j^{(hn_0)} \;\leq\; \big(M_j^{(0)} - m_j^{(0)}\big)(1-\varepsilon)^h \;=\; (1-\varepsilon)^h, \qquad \forall\, j \in \mathcal{S},$$

dove l'ultima uguaglianza segue da $M_j^{(0)} = 1$ e $m_j^{(0)} = 0$. □

Il Teorema 4.1 fornisce un collegamento tra π_j e la probabilità di transizione $\pi_{ij}^{(n)}$ (che corrisponde alla probabilità condizionata $\mathbb{P}(X_{n+1} = j|X_1 = i)$ o, più in generale, a $\mathbb{P}(X_{n+k} = j|X_k = i)$).

Più precisamente, il Teorema 4.1 ci dice che π_j è approssimativamente uguale a $\pi_{ij}^{(n)}$ per $n \gg 1$, qualunque sia i. Ciò fornisce la seguente interpretazione di π_j: la probabilità π_j rappresenta, approssimativamente, la probabilità (condizionata) $\mathbb{P}(X_{n+1} = j|X_1 = i)$ di essere nello stato j al passo n, indipendentemente da quale sia lo stato di partenza i.

Di conseguenza, gli stati j per cui la probabilità π_j è elevata sono quelli più facilmente raggiungibili dagli altri stati. In altri termini, sono gli stati che vengono più spesso "visitati" se si immagina di partire da un qualunque stato i e di percorrere una passeggiata aleatoria lungo il grafo orientato associato alla catena di Markov.

Ordinando gli stati in modo crescente in base ai valori contenuti in $\vec{\boldsymbol{\pi}}$, si ottiene quindi un ordinamento degli stati in base a quanto sono più o meno facilmente *"visitabili"*.

4.2 Google e l'algoritmo PageRank

Vediamo ora un'applicazione delle catene di Markov di grande successo: l'algoritmo PageRank utilizzato dal motore di ricerca Google, ideato nel 1997 da Sergey Brin e Lawrence "Larry" Page. Quanto segue si basa sul seguente articolo di ricerca:

S. Brin, L. Page. *The anatomy of a large-scale hypertextual web searching engine.* Computer networks and ISDN Systems, 33, 107-117, 1998.

Come spiegato in questo articolo, 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, intesa come grafo orientato in cui ogni nodo corrisponde ad una pagina e ogni freccia rappresenta un link. Tale struttura permette di ordinare le pagine, ovvero 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'algoritmo PageRank nella sua forma attuale non è chiaramente noto (essendo ovviamente tutelato da copyright), però l'idea originale del metodo è chiara. 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 $\mathrm{PR}(A)$. 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;

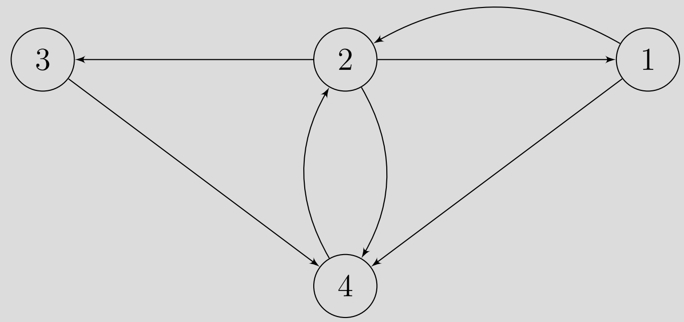
2) risultare elevato se riferito ad una pagina citata da (eventualmente poche) pagine molto significative.

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.

4.3 Descrizione dell'algoritmo PageRank

Rappresentiamo graficamente il web come un grafo orientato in cui ogni nodo corrisponde ad una pagina e ogni freccia ad un link. Immaginiamo 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 $(X_n)_n$ la catena di Markov che descrive tale passeggiata aleatoria e con Π la sua matrice di transizione. Vediamo a tal proposito il seguente esercizio.

Esercizio 4.1. *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* $(X_n)_n$ *la catena di Markov che descrive tale passeggiata aleatoria.*

(a) *Qual è la distribuzione iniziale della catena di Markov, ovvero qual è la densità discreta di* X_1*?*

(b) *Si rappresenti graficamente la catena di Markov tramite un grafo orientato.*

(c) *Qual è la matrice di transizione di* $(X_n)_n$*?*

Nella realtà, la dimensione della matrice Π è gigantesca; si pensi che già nel 1997 si stimava la presenza di circa 100 milioni di pagine web.

Si noti 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, il loro valore è dato da $1/m_i$ (per l'ipotesi di equiprobabilità, tutti hanno lo stesso valore). 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 PageRank è quella di considerare la *distribuzione invariante* $\vec{\boldsymbol{\pi}}$ della catena di Markov, ordinando gli stati utilizzando proprio l'ordinamento suggerito da $\vec{\boldsymbol{\pi}}$, quindi ponendo

$$\mathrm{PR}(j) \; = \; \pi_j.$$

per ogni pagina web j. Vediamo con un esempio perché π_j è un buon indice di significatività (ovvero verifica i due requisiti precedentemente riportati sopra). Per farlo riprendiamo l'Esercizio 4.1 e determiniamo in tal caso la distribuzione invariante $\vec{\boldsymbol{\pi}}$.

Esercizio 4.1 (continuazione). *Dobbiamo determinare* $\vec{\boldsymbol{\pi}}$. *Dato che*

$$\vec{\boldsymbol{\pi}} \;=\; \vec{\boldsymbol{\pi}}\,\Pi$$

e

$$\Pi \;=\; \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/3 & 0 & 1/3 & 1/3 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix},$$

si ottiene il sistema di equazioni seguente:

$$\begin{cases} \pi_1 &= \frac{1}{3}\pi_2, \\ \pi_2 &= \frac{1}{2}\pi_1 + \pi_4, \\ \pi_3 &= \frac{1}{3}\pi_2, \\ \pi_4 &= \frac{1}{2}\pi_1 + \frac{1}{3}\pi_2 + \pi_3. \end{cases}$$

Si noti che $\vec{\boldsymbol{\pi}}$ *è un vettore densità discreta, quindi deve verificare anche le due proprietà seguenti:*

1) $0 \leq \pi_j \leq 1$, *per ogni* $j = 1, 2, 3, 4$;

2) $\sum_{j=1}^{4} \pi_j = 1$.

Si ottiene dunque un'unica soluzione al sistema precedente, che è data da:

$$\begin{cases} \pi_1 &= 0.133333, \\ \pi_2 &= 0.399999, \\ \pi_3 &= 0.133333, \\ \pi_4 &= 0.333333. \end{cases}$$

L'ordinamento delle pagine ottenuto in tal modo è: $\mathrm{Pag}_2 > \mathrm{Pag}_4 > \mathrm{Pag}_1 = \mathrm{Pag}_3$.

Osservazione. *Dal sistema di equazioni dell'esempio precedente si deduce che l'indice di significatività di una pagina web* A, *a cui si accede dalle pagine* $T_1, \ldots, T_n$, *è dato da*

$$\mathrm{PR}(A) \;=\; \frac{1}{\mathrm{C}(T_1)}\mathrm{PR}(T_1) + \cdots + \frac{1}{\mathrm{C}(T_n)}\mathrm{PR}(T_n),$$

dove $\mathrm{C}(T_i)$ *è il numero di link che partono dalla pagina* T_i. *In conclusione, come richiesto, l'indice di significatività di* A *è proporzionale agli indici di significatività delle pagine che conducono ad* A.

Concludiamo infine osservando che quanto detto finora si basa sul Teorema 4.1, che però vale solo quando $(X_n)_n$ è 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 come si affronta questo problema.

Nel web ci sono 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.

Per risolvere questo problema una possibilità è quella di ipotizzare che ciascuna pagina dangling punti ad ogni altra pagina del web (ciò è giustificato dal fatto che si passa da una pagina all'altra non solo tramite i link, ma anche con la barra degli indirizzi): questo equivale a sostituire alla riga di una pagina dangling, formata da tutti zeri, il vettore riga

$$\begin{bmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{bmatrix}.$$

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

Nonostante questa modifica la catena di Markov associata a $\widetilde{\Pi}$ non è ancora *regolare* (quindi non è ancora possibile applicare il Teorema ergodico). Per ottenere tale proprietà è sufficiente perturbare la matrice $\widetilde{\Pi}$ come segue:

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

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 $\widetilde{\Pi}$, quindi è facile mostrare che anche 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 4.1.