APPUNTI SULLE CATENE DI MARKOV

DARIO TREVISAN

Indice

1. Processi stocastici

Spesso abbiamo a che fare con "quantità" (non necessariamente numeri) incerte che cambiano nel tempo. Nel linguaggio matematico sono rappresentate da *famiglie di variabili aleatorie* indicizzate mediante un parametro "tempo".

Esempio 1. Nel modello delle estrazioni dall'urna (contenente palline rosse o blu) possiamo considerare il colore della pallina alla prima, seconda, terza ecc. estrazione. Questa "quantità" che varia nel "tempo" (dato dall'ordine dell'estrazione) è rappresentabile con la famiglia di variabili aleatorie

$$X_1, X_2, \dots, X_n \in \{\text{rossa}, \text{blu}\}.$$

se $n \geq 1$ è il numero totale di estrazioni che vengono effettuate.

Esempio 2. La temperatura della CPU di un certo computer è incerta (ad esempio se siamo in fase di progettazione del computer) e inoltre varia nel tempo – a seconda della frequenza, dell'utilizzo ecc.

Un modello potrebbe essere dato da una famiglia di variabili aleatorie

$$X_{t_0}, X_{t_1}, X_{t_2}, \dots, X_{t_n} \in \mathbb{R}$$

Appunti del corso CPS 269AA A.A. 2017-2018, CdL in Informatica. Vi prego di segnalare errori di battitura, punti poco chiari ecc., via e-mail a dario.trevisan@unipi.it.

dove $t_0 < t_1 < \ldots < t_n$ sono $n+1$ istanti di tempo, e la temperatura assume valori reali (ovviamente molto dipende dal grado di precisione del termometro, adesempio potrebbero essere numeri interi).

Esercizio 3. Provate a modelizzare come negli esempi sopra altre situazioni realistiche di "quantità" incerte che cambiato nel "tempo", associando ad esse una famiglia di variabili aleatorie: considerate ad esempio il numero di email nella casella di posta elettronica di una persona, il numero di persone che salgono su i treni tra Pisa e Firenze, le intenzioni di voto di una "generica" persona, il colore che sarà più in voga nell'abbigliamento dei prossimi anni, ecc. . .

Definizione 4 (Processo stocastico)**.** Una famiglia di variabili aleatorie $\{X_t\}_{t\in\mathcal{T}}$, dove $\mathcal{T} \subseteq \mathbb{R}$ e che assumono tutte valori nello stesso insieme E è detta *processo stocastico*. L'insieme $\mathcal{T}$ è detto insieme dei *tempi* del processo e l'insieme E è detto insieme (o spazio) degli *stati* del processo.

Osservazione 5. L'insieme degli stati potrebbe essere anche molto grande (pensate all'esempio della temperatura), ma in seguito ci ridurremo *sempre* ad un insieme di stati *discreto* (numerabile) o addirittura *finito*, ossia di lavorare con variabili aleatorie discrete. Valutate negli esempi e nell'esercizio precedenti se è possibile (o ragionevole) ridursi ad un insieme finito di stati E per i vari processi.

Osservazione 6. Allo stesso modo, l'insieme dei tempi $\mathcal{T}$ potrebbe essere un intervallo $\mathcal{T} = [0, T]$, e spesso idealmente si vorrebbe proprio questo, però ci sono diverse complicazioni matematiche. Per questo noi considereremo solamente un insieme di tempi *discreto* (numerabile, ad esempio $\mathcal{T} = \mathbb{N}$ o $\mathcal{T} = \mathbb{Z}$) o anche più spesso *finito* ($\mathcal{T} = \{0, 1, \ldots, n\}$ o più in generale $\mathcal{T} = \{t_0, t_1, \ldots, t_n\}$, come nell'Esempio 2).

Definizione 7. Dato un processo $\{X_t\}_{t\in\mathcal{T}}$, le variabili aleatorie $X_t \in E$ sono dette *marginali* del processo.

Spesso le marginali sono anche dette 1-marginali, per poi chiamare 2-marginali le variabili (X_t, X_s) date da una qualunque coppia di tempi s, $t \in \mathcal{T}$, ecc.

Osservazione 8. Notate che le leggi (ad esempio rispetto all'informazione iniziale Ω) delle marginali di due processi potrebbero coincidere, pur essendo i due processi molto diversi: nell'Esempio 1, se le estrazioni sono con re-immissione o senza, i due processi cambiano radicalmente, però sappiamo che le marginali hanno tutte le stesse leggi (rispetto a $P(\cdot|\Omega)$)

$$P(X_k = \text{rossa}|\Omega) = \frac{\sharp\text{palline rosse}}{\sharp\text{palline totali}}.$$

Lo studio dei processi stocastici comprende in particolare la capacità di fare *inferenze* (previsioni probabilistiche) basandoci sull'informazione di cui disponiamo. Un esempio naturale è quello di fare previsioni sui tempi "futuri" basandoci sull'informazione delle osservazioni passate (e magari qualche ulteriore informazione), però potremmo essere interessati a fare inferenza sui tempi "passati": pensate al lavoro di un archeologo o ad un detective.

I processi più semplici in cui modelizziamo tale capacità sono i processi di Markov. Vediamo prima per contrasto un caso molto semplice in cui non riusciamo ad inferire nulla dal passato (sarà comunque un processo di Markov, ma abbastanza speciale).

Esempio 9. Consideriamo l'Esempio 1 in cui le estrazioni sono *con* reimmissione, situazione che abbiamo modellizzato assumendo che le variabili $X_1, \ldots, X_n \in \{\text{rossa}, \text{blu}\}$ siano tutte *indipendenti* (rispetto all'informazione iniziale Ω, che comprende anche il numero di palline totali e il numero di palline rosse, che quindi conosciamo). Supponiamo di avere fatto $k < n$ estrazioni e di conoscere esattamente il loro esito, ad esempio poniamo che siano tutte rosse:

$$X_1 = \text{rossa}, X_2 = \text{rossa}, \ldots, X_k = \text{rossa}.$$

Chiediamoci quindi, assumendo questa nuova informazione, quale sia la probabilità che all'estrazione $k+1$ otteniamo una pallina rossa:

$$\begin{aligned} &P(X_{k+1} = rossa | \Omega \cap \{X_1 = \text{rossa}, X_2 = \text{rossa}, \ldots, X_k = \text{rossa}\}) \\ &\quad = P\left(X_{k+1} = rossa | \Omega\right) \quad \text{(per indipendenza)} \\ &\quad = \frac{\sharp \text{palline rosse}}{\sharp \text{palline totali}}. \end{aligned}$$

L'ipotesi di indipendenza probabilistica si traduce quindi nel fatto che non siamo capaci di "imparare" dal passato.

2. Proprietà di Markov

I processi di Markov sono tali per cui l'informazione ottenibile dall'osservazione della storia del processo fino al presente può essere utile a fare inferenza sullo stato futuro. In realtà costituiscono la classe più semplice, in cui tutta la storia passata, eccetto il presente, può essere trascurata ai fini dell'inferenza sul futuro. In parole più evocative (ma comunque corrette):

un processo è di Markov se,
conoscendo (esattamente) il presente,
passato e futuro sono indipendenti.

Diamo la definizione matematica della proprietà di Markov per un processo a tempi $\mathcal{T}$ finiti e stati E discreti.

Definizione 10. Un processo $\{X_{t_i}\}_{i=0,\ldots,n}$ con $t_0 < t_1 < \ldots < t_n$ è detto *di Markov* (rispetto a $P(\cdot|I)$) se, per ogni $k \in \{1, \ldots, n-1\}$, e_0, e_1, $\ldots$, $e_{k+1} \in E$, vale

$$\begin{aligned} &P\left(X_{t_{k+1}} = e_{k+1} | I \cap \{X_{t_k} = e_k\} \cap \left\{X_{t_{k-1}} = e_{k-1}\right\} \cap \ldots \{X_{t_0} = e_0\}\right) = \\ &\quad = P\left(X_{t_{k+1}} = e_{k+1} | I \cap \{X_{t_k} = e_k\}\right) \end{aligned}$$

Ovviamente la proprietà di Markov non riguarda quasi mai la realtà "fisica" del modello che cerchiamo di studiare, si tratta invece di una ipotesi molto semplificativa (ma non troppo) che permette di svolgere calcoli e fare predizioni. Infatti, in generale, l'indipendenza probabilistica va sempre considerata come una ipotesi che noi introduciamo nel modello, perché non siamo capaci (meglio: non abbiamo ragione di credere) di sfruttare una certa informazione.

Osservazione 11. Ricordando la definizione di variabili aleatorie indipendenti, si potrebbe anche dire più semplicemente che rispetto a tutte le informazioni della forma $\{X_{t_k} = e_k\}$, le variabili $X_{t_{k+1}}$, $X_{t_{k-1}}$ $\ldots X_{t_0}$ sono indipendenti tra loro.

Esempio 12. Torniamo agli esempi della sezione precedente. Nel caso di estrazioni di un'urna di cui conosciamo esattamente il contenuto, se le estrazioni sono con reimmissione, il processo è sicuramente di Markov. Invece, se le estrazioni sono senza reimmissione, il processo non è di Markov: tutta la sequenza di palline estratte è necessaria per conoscere il contenuto esatto dell'urna, quindi l'informazione passata non può essere trascurata (pur mantenendo l'informazione dell'ultima pallina estratta).

Esempio 13. Un esempio di processo di Markov basato su estrazioni dall'urna è il seguente. Si consideri un urna contenente R palline rosse e B palline blu (in tutto contiene $N = R + B$ palline). Una persona svolge in sequenza estrazioni nel seguente modo: inizialmente estrae una pallina, e la tiene con sé, poi ne estrae un'altra e successivamente rimette dentro l'urna la prima pallina estratta (e agita l'urna). Procede poi con ulteriori estrazioni, tenendo sempre fuori dall'urna solamente l'ultima pallina estratta. In questo caso, ponendo X_k il colore della pallina estratta all'estrazione k, la proprietà di Markov vale (infatti l'informazione di tutta la sequenza di estrazioni non è rilevante eccetto l'ultima, che è la pallina rimasta fuori).

Dati i, $j \subseteq E$, le *probabilità di transizione*

$$P\left(X_{t_{k+1}} = j | I \cap X_{t_k} = i\right)$$

potrebbero in generale dipendere da k. Studieremo il caso in cui queste non dipendono da k (in cui i calcoli sono notevolmente semplificati). Per semplificare ulteriormente, assumiamo che $t_k = k$ in seguito.

Definizione 14. Un processo di Markov $\{X_i\}_{i=0,\ldots,n}$ è *omogeneo* se le probabilità di transizione non dipendono da $k \in \{0, 1, \ldots, n-1\}$, ossia

$$P\left(X_{k+1} = j | I \cap \{X_k = i\}\right) = P\left(X_1 = j | I \cap \{X_0 = i\}\right) \quad \text{per ogni } i,\, j \in E.$$

Questa definizione ci permette di collezionare le probabilità di transizione in una singola *matrice di transizione* (quadrata con $\sharp E$ righe e $\sharp E$ colonne):

$$Q_{ij} = Q_{i \to j} := P\left(X_1 = j | I \cap \{X_0 = i\}\right), \quad \text{per } i,j \in E.$$

Osservazione 15. Per scrivere in pratica una matrice di transizione Q bisogna fissare un *ordinamento* degli stati E (ossia decidere quale stato è rappresentato nella prima riga, nella seconda, ecc.). Una volta fissato questo ordinamento (vi consiglio di scriverlo esplicitamente negli esercizi) seguitelo consistentemente in tutto il problema.

Osservazione 16. Le entrate della matrice di transizione $Q_{i\to j}$ sono tutti numeri compresi tra 0 e 1 (sono probabilità) e inoltre, per ogni $i \in E$,

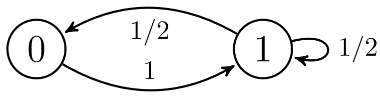
$$\sum_{j \in E} Q_{i \to j} = \sum_{j \in E} P\left(X_1 = j | I \cap \{X_0 = i\}\right) = P\left(X_1 \in E | I \cap \{X_0 = i\}\right) = 1,$$

ossia la somma delle entrate in ciascuna riga vale 1.

Definizione 17. Un processo di Markov omogeneo $\{X_i\}_{i=0,\dots,n}$ a stati finiti (o discreti) è detto *catena di Markov* (omogenea).

Per *visualizzare* una catena di Markov con una certa matrice di transizione, può essere utile associarle un *grafo* orientato, costruito nel seguente modo: ad ogni stato $i \in E$, facciamo corrispondere un nodo, e ad ogni probabilità di transizione $Q_{i\to j}$ strettamente positiva, disegniamo un arco orientato (una freccia) dal nodo i al nodo j, su cui segniamo anche la probabilità $Q_{i\to j}$. È utile *non* disegnare le frecce corrispondenti a probabilità di transizione nulle, perché non daranno alcun contributo ai calcoli. Notate però che la questa rappresentazione non dice nulla sulle leggi marginali della catena, in particolare sulla legge iniziale di X_0.

Esempio 18. Rappresentiamo la catena su $E = \{0, 1\}$ con probabilità di transizione $Q_{0\to1} = 1$, $Q_{0\to0} = 0$, $Q_{1\to0} = Q_{1\to1} = 1/2$:

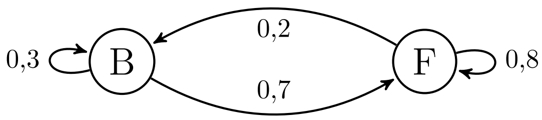


3. Esempi

Esempio 19. Consideriamo una risorsa (ad es. una CPU) che può trovarsi in due stati: “occupato” (BUSY) oppure “libero” (FREE). Un progettista vuole rappresentare lo stato della risorsa mediante una catena di Markov in cui $E = \{\mathrm{BUSY}, \mathrm{FREE}\}$ e la matrice di transizione è

$$Q = \begin{pmatrix} 0,3 & 0,7 \\ 0,2 & 0,8 \end{pmatrix}.$$

(dove scriviamo nella prima riga/colonna lo stato BUSY e nella seconda lo stato FREE). Diamone la rappresentazione grafica (per brevità scriviamo B, F invece di BUSY e FREE)



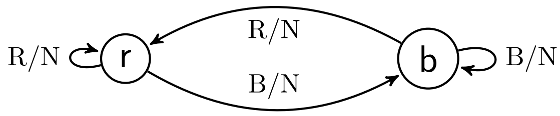
Esempio 20. Consideriamo la situazione di un’urna contenente R palline rosse e B palline blu (in tutto $N = R + B$ palline) da cui una persona effettua estrazioni con reimmissione. Abbiamo già visto (Esempio 9) che la proprietà di Markov vale. Le probabilità di transizione sono molto semplici da calcolare, grazie all’indipendenza delle variabili $\{X_k\}$:

$$P(X_{k+1} = \text{rossa} | \Omega \cap \{X_k = \text{rossa}\}) = P(X_{k+1} = \text{rossa}|\Omega) = \frac{R}{N}$$
$$P(X_{k+1} = \text{blu} | \Omega \cap \{X_k = \text{rossa}\}) = P(X_{k+1} = \text{blu}|\Omega) = \frac{B}{N}$$
$$P(X_{k+1} = \text{rossa} | \Omega \cap \{X_k = \text{blu}\}) = P(X_{k+1} = \text{rossa}|\Omega) = \frac{R}{N}$$
$$P(X_{k+1} = \text{blu} | \Omega \cap \{X_k = \text{blu}\}) = P(X_{k+1} = \text{blu}|\Omega) = \frac{B}{N}.$$

Quindi la catena di Markov sugli stati $E = \{\text{rossa}, \text{blu}\}$ è omogenea con matrice di transizione (dove scriviamo nella prima riga/colonna lo stato "rossa", e sulla seconda "blu")

$$Q = \begin{pmatrix} R/N & B/N \\ R/N & B/N \end{pmatrix}.$$

Diamone la rappresentazione grafica (per brevità scriviamo r, b invece di "rossa" e "blu")



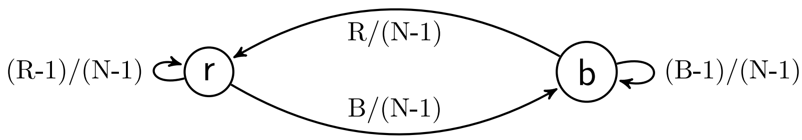
Esempio 21. Consideriamo la situazione di un'urna contenente $R \geq 1$ palline rosse e $B \geq 1$ palline blu (in tutto $N = R + B$ palline) da cui una persona effettua estrazioni con la modalità descritta nell'Esempio 13. Calcoliamo le probabilità di transizione

$$P\left(X_{k+1} = \text{rossa} | \Omega \cap \{X_k = \text{rossa}\}\right) = P\left(X_{k+1} = \text{rossa} | \text{tolta una rossa}\right) = \frac{R-1}{N-1}$$
$$P\left(X_{k+1} = \text{blu} | \Omega \cap \{X_k = \text{rossa}\}\right) = P\left(X_{k+1} = \text{blu} | \text{tolta una rossa}\right) = \frac{B}{N-1}$$
$$P\left(X_{k+1} = \text{rossa} | \Omega \cap \{X_k = \text{blu}\}\right) = P\left(X_{k+1} = \text{rossa} | \text{tolta una blu}\right) = \frac{R}{N-1}$$
$$P\left(X_{k+1} = \text{blu} | \Omega \cap \{X_k = \text{blu}\}\right) = P\left(X_{k+1} = \text{blu} | \text{tolta una blu}\right) = \frac{B-1}{N-1}.$$

Quindi la catena di Markov sugli stati $E = \{\text{rossa}, \text{blu}\}$ è pure omogenea, ma la matrice di transizione è diversa dall'esempio precedente (scriviamo nella prima riga/colonna lo stato "rossa", e sulla seconda "blu")

$$Q = \begin{pmatrix} (R-1)/(N-1) & B/(N-1) \\ R/(N-1) & (B-1)/(N-1) \end{pmatrix}.$$

Diamone la rappresentazione grafica (per brevità scriviamo r, b invece di "rossa" e "blu")



Esempio 22 (Passeggiata aleatoria simmetrica). Consideriamo una successione $\{Y_k\}_{k=1,\ldots,n}$ di variabili aleatorie indipendenti con $P(Y_k = 1|\Omega) = P(Y_k = -1|\Omega)1/2$, per ogni $k \in \{1, \ldots, n\}$. Definiamo la "passeggiata aleatoria" $\{X_k\}_{k=0,\ldots,n}$ nel seguente modo (ricorsivo)

$$\begin{cases} X_0 = 0 \\ X_k = X_{k-1} + Y_k \quad \text{per } k \in \{1, \ldots, n\}. \end{cases}$$

Possiamo darne una descrizione intuitiva, nel seguente modo: una quantità X è inizialmente 0 ($X_0 = 0$). Ad ogni tempo k, questa quantità sale di 1 o scende di 1 rispetto all'istante precedente, con incertezza uniforme (dal punto di vista di chi possiede l'informazione iniziale Ω). Notiamo che l'insieme degli stati è $E = \{-n, -(n-1), \ldots, -1, 0, 1, \ldots, n-1, n\}$. A volte è più semplice considerare direttamente su $E = \mathbb{N}$, anche se è infinito, perché così non ci sono problemi quando il processo raggiunge il "bordo" $\{-n, n\}$.

Il processo è di Markov, perché conoscere tutte le posizioni fino al tempo k o conoscere solamente la posizione al tempo k influenzano allo stesso modo la probabilità che il processo al tempo $k+1$ si trovi in una certa posizione:

$$\begin{aligned} &P(X_{k+1} = j | \Omega \cap \{X_k = i\} \cap \ldots \cap \{X_0 = i_0\}) = \\ &= P(X_k + Y_{k+1} = j | \Omega \cap \{X_k = i\} \cap \ldots \cap \{X_0 = i_0\}) \\ &= P(i + Y_{k+1} = j | \Omega \cap \{X_k = i\} \cap \ldots \cap \{X_0 = i_0\}) \\ &= P(i + Y_{k+1} = j | \Omega) \end{aligned}$$

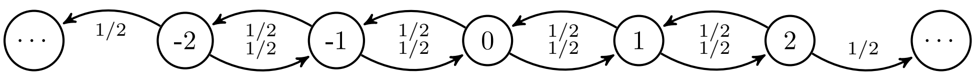
dove abbiamo usato l'indipendenza delle $(Y_\ell)_{\ell=1,\ldots,n}$: infatti l'informazione rispetto alla quale stiamo condizionando si può esprimere usando solamente le Y_ℓ con $\ell \in \{1, \ldots, k\}$, e quindi si può trascurare quando calcoliamo la probabilità di un evento che riguarda solamente Y_{k+1}. Per lo stesso motivo (indipendenza), possiamo scrivere

$$\begin{aligned} P(i + Y_{k+1} = j | \Omega) &= P(i + Y_{k+1} = j | \Omega \cap \{X_k = i\}) \\ &= P(X_k + Y_{k+1} = j | \Omega \cap \{X_k = i\}) \\ &= P(X_{k+1} = j | \Omega \cap \{X_k = i\}) \end{aligned}$$

che dimostra la proprietà di Markov (anche se intuitivamente era chiara). Inoltre, questa identità ci permette di calcolare le probabilità di transizione (che pure intuitivamente sono chiare):

$$P(X_{k+1} = j | \Omega \cap \{X_k = i\}) = \begin{cases} 1/2 & \text{se } j = i+1 \\ 1/2 & \text{se } j = i-1 \\ 0 & \text{altrimenti.} \end{cases}$$

La catena di Markov è quindi omogenea: diamone la rappresentazione grafica (parziale, per ragioni di spazio...):



4. Algebra lineare e catene di Markov

L'identificazione delle probabilità di transizione con una matrice $Q = (Q_{i \to j})_{i,j \in E}$ è molto comoda perché permette di ridurre alcuni calcoli a semplici operazioni tra matrici e vettori (il cui svolgimento possiamo affidare ad un computer, ad esempio). In questa sezione solamente il problema del calcolo delle leggi marginali e dei valori attesi. Notate però che non tutti i problemi si riducono a questi due, ed è importante ricordare che le proprietà fondamentali delle catene di Markov sono la proprietà di Markov e l'omogeneita (e queste possono servire per risolvere più problemi).

4.1. **Calcolo di leggi marginali.** Data una catena di Markov $(X_k)_{k=0,\ldots,n}$, supponiamo di conoscere la legge iniziale

$$P(X_0 = i | \Omega), \quad \text{per } i \in E.$$

e la matrice di transizione Q. Come possiamo calcolare la legge della marginale al tempo $k \geq 1$, ossia $P(X_k = j|\Omega)$, $j \in E$?

La risposta è algebricamente molto semplice: associamo alla legge iniziale un vettore *riga*

$$\vec{\mu}_i := P(X_0 = i|\Omega)$$

mantenendo lo stesso ordinamento degli stati E che usiamo nella matrice di transizione. A questo punto basta calcolare il prodotto

$$\vec{\mu} \cdot Q^k$$

dove $Q^k = Q \cdot Q \cdot \ldots \cdot Q$ è il prodotto (righe per colonne) della matrice Q con se stessa. Il risultato dell'operazione sopra è ancora un vettore riga, nelle cui componenti leggiamo (sempre nell'ordine) le probabilità richieste:

$$P(X_k = j|\Omega) = \left(\vec{\mu} \cdot Q^k\right)_i \quad i \in E.$$

Diamone una dimostrazione nel caso $k = 1$ (il caso generale segue per induzione). Decomponendo con il sitema di alternative $\{X_0 = i\}$, $i \in E$,

$$\begin{aligned} P(X_1 = j|\Omega) &= \sum_{i \in E} P(X_1 = j|\Omega \cap (X_0 = i))P(X_0 = i|\Omega) \\ &= \sum_{i \in E} Q_{i \to j} \vec{\mu}_i = (\vec{\mu} \cdot Q)_j \,. \end{aligned}$$

Osservazione 23. Scegliendo $\vec{\mu} = e_i$ i vettori riga della base canonica (tutti zero eccetto nella posizione i, in cui abbiamo 1) ne otteniamo in particolare che la matrice Q^k nella riga i e alla colonna j vale esattamente

$$(Q^k)_{ij} = P\left(X_k = j|\Omega \cap \{X_0 = i\}\right),$$

quindi anche le entrate di Q^k sono probabilità (e le somme sulle righe valgono tutte 1).

4.2. **Calcolo di valori attesi.** Un problema simile (matematicamente detto *duale*) è il seguente: data una catena di Markov $(X_k)_{k=0,\ldots,n}$, e una funzione

$$f : E \mapsto \mathbb{R}$$

calcolare il valore atteso $\mathbb{E}\left[f(X_k)|\Omega \cap \{X_0 = i\}\right]$, per $i \in E$. Anche stavolta la risposta è algebricamente molto semplice: associamo alla funzione f un vettore *colonna*

$$\vec{f}_j := f(j)$$

mantenendo lo stesso ordinamento degli stati E che usiamo nella matrice di transizione. A questo punto basta calcolare il prodotto

$$Q^k \cdot \vec{f}$$

che è un vettore colonna: la componente $i \in E$ corrisponde alla quantità richiesta

$$\mathbb{E}\left[f(X_k)|\Omega \cap \{X_0 = i\}\right] = \left(Q^k \cdot \vec{f}\right)_i \quad \text{per } i \in E.$$

Anche in questo caso una dimostrazione generale si potrebbe fare per induzione (oppure usando il risultato precedente), per semplicità ragioniamo nel caso $k = 1$, usando il sistema di alternative $\{X_1 = j\}$, $j \in E$,

$$\begin{aligned}
&\mathbb{E}\left[f(X_1)|\Omega \cap \{X_0 = i\}\right] = \\
&\quad = \sum_{j \in E} \mathbb{E}\left[f(X_1)|\Omega \cap \{X_0 = i\} \cap \{X_1 = j\}\right] P(X_1 = j|\Omega \cap (X_0 = i)) \\
&\quad = \sum_{j \in E} \mathbb{E}\left[f(j)|\Omega \cap \{X_0 = i\} \cap \{X_1 = j\}\right] Q_{i \to j} \\
&\quad = \sum_{j \in E} \vec{f}_j Q_{i \to j} = \left(Q \cdot \vec{f}\right)_i .
\end{aligned}$$

5. Catene di Markov stazionarie

Un concetto importante è quello di studio di un sistema all'*equilibrio*. Intuitivamente questo significa che le sue condizioni non cambiano nel tempo, e (idealmente) questo si raggiunge isolando il sistema e attendendo abbastanza a lungo. Ovviamente queste condizioni ideali non esistono per molti fenomeni, tuttavia spesso forniscono un modello molto utile.

Lo studio delle catene di Markov all'equilibrio e della loro convergenza occupa una gran parte della teoria sull'argomento (anche con risultati soprendenti: ad esempio il matematico P. Diaconis servendosi anche di questa teoria consiglia ai tavoli di blackjack di mescolare 7 volte il mazzo di carte, per distribuirle "bene"!). Noi diamo solamente due definizioni fondamentali, e supponiamo che E sia un insieme finito (per semplicità).

Definizione 24 (distribuzione invariante)**.** Un vettore riga $\vec{\mu} = (\vec{\mu}_i)_{i \in E}$ è detto *distribuzione invariante* per una matrice di transizione Q (di una catena di Markov omogenea) se vale

$$\vec{\mu}_i \geq 0, \quad \sum_{i \in E} \vec{\mu}_i = 1, \quad \text{e} \quad \vec{\mu} \cdot Q = \vec{\mu}.$$

Come si vede, la definizione è completamente *algebrica* e in particolare l'ultima significa che il vettore colonna $\vec{\mu}^\tau$ (τ sta per l'operazione di trasposizione) è un autovettore di autovalore 1 per la matrice Q^τ (la matrice trasposta). Quindi per trovare una distribuzione invariante bisogna intanto risolvere il sistema $(Q^\tau - \mathrm{Id})\vec{v} = 0$ (Id è la matrice identità).

Definizione 25 (catena stazionaria)**.** Una catena di Markov $(X_k)_{k=0,1,\ldots,n}$ (rispetto all'informazione I) è detta stazionaria se tutte le sue leggi marginali sono uguali, ossia

$$P(X_k = i|I) = P(X_0 = i|I), \quad \text{per } k \in \{1, \ldots, n\},\ i \in E.$$

Osservazione 26. Grazie ai risultati della sezione precedente (calcolo di leggi marginali), otteniamo che, posto $\vec{\mu}$ il vettore riga corrispondente alla legge di X_0, la catena $(X_k)_{k=0,\ldots,n}$ è stazionaria se (e solo se)

$$\vec{\mu} \cdot Q^k = \vec{\mu}, \quad \text{per ogni } k \in \{1, \ldots, n\},$$

ma questa condizione basta verificarla per $k = 1$, così

$$\vec{\mu} \cdot Q^k = \bar{\mu} \cdot Q \cdot Q^{k-1} = \vec{\mu} \cdot Q^{k-1} = \ldots = \vec{\mu}.$$

e quindi equivale al fatto che il vettore $\vec{\mu}$ sia una distribuzione invariante.

Esempio 27. Consideriamo la matrice di transizione Q dell'Esempio 2, e calcoliamo

$$Q^\tau - \mathrm{Id} = \begin{pmatrix} 0,3-1 & 0,2 \\ 0,7 & 0,8-1 \end{pmatrix} = \begin{pmatrix} -0,7 & 0,2 \\ 0,7 & -0,2 \end{pmatrix}.$$

Sommando la prima riga alla seconda, otteniamo la matrice (equivalente ai fini della soluzione del sistema omogeneo $(Q^\tau - \mathrm{Id})\vec{v} = 0$)

$$\begin{pmatrix} -0,7 & 0,2 \\ 0 & 0 \end{pmatrix}$$

da cui deduciamo che tutti gli autovettori (colonna) di autovalore 1 sono della forma $t(2/7, 1)$, $t \neq 0$. Se cerchiamo le distribuzioni invarianti, deve valere inoltre

$$t\frac{2}{7} + t = 1 \Rightarrow t = \frac{7}{9},$$

da cui l'unica distribuzione invariante è $\bar{\mu} = (2/9, 7, 9)$. Pertanto una catena di Markov con matrice di transizione Q è stazionaria se e solo se

$$P(X_0 = \mathrm{BUSY}|\Omega) = \frac{2}{9}, \quad P(X_0 = \mathrm{FREE}|\Omega) = \frac{7}{9}.$$

In termini di *equilibrio*, possiamo dire che all'equilibrio la risorsa è occupata con probabilità 7/9 (e libera con probabilità 2/9). Osservando le probabilità di transizione, questo è consistente con il fatto che, se la macchina è occupata, resta occupata con alta probabilità $(0,8)$.

Esempio 28. Consideriamo la matrice di transizione Q dell'Esempio 20, e calcoliamo

$$Q^\tau - \mathrm{Id} = \begin{pmatrix} R/N-1 & R/N \\ B/N & B/N-1 \end{pmatrix} = \begin{pmatrix} -B/N & R/N \\ B/N & -R/N \end{pmatrix}.$$

Con operazioni di riga, otteniamo la matrice equivalente

$$\begin{pmatrix} -B & R \\ 0 & 0 \end{pmatrix}$$

da cui gli autovettori con autovalore 1 sono della forma $t(R, B)$ $(t \neq 0)$, e dovendo sommare a 1, otteniamo che l'unica distribuzione invariante è

$$\vec{\mu} = \left(\frac{R}{N}, \frac{B}{N}\right).$$

C'era qualche altro modo per arrivare a questa soluzione?

Esempio 29. Consideriamo la matrice di transizione Q dell'Esempio 21, e calcoliamo

$$\begin{aligned} Q^\tau - \mathrm{Id}\, Q &= \begin{pmatrix} (R-1)/(N-1)-1 & R/(N-1) \\ B/(N-1) & (B-1)/(N-1)-1 \end{pmatrix} \\ &= \begin{pmatrix} -B/(N-1) & R/(N-1) \\ B/(N-1) & -R/(N-1) \end{pmatrix}. \end{aligned}$$

Con operazioni di riga torniamo alla matrice equivalente

$$\begin{pmatrix} -B & R \\ 0 & 0 \end{pmatrix}$$

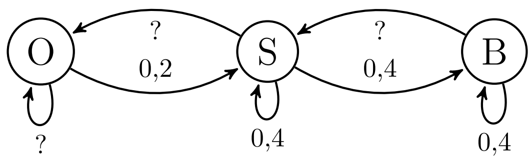
dell'esempio precedente, quindi anche in questo caso l'unica distribuzione invariante è

$$\vec{\mu} = \left(\frac{R}{N}, \frac{B}{N} \right).$$

Il calcolo mostra quindi che l'effetto di togliere l'ultima (e solo l'ultima) pallina estratta non influenza molto la distribuzione all'*equilibrio* della catena.

6. Problemi

Problema 30. Una macchina può trovarsi in tre differenti stati: spenta (OFF, O), stand-by (S), oppure in uso (BUSY, B). Un progettista rappresenta il suo stato mediante la catena di Markov associata alla figura:



(1) Completare le probabilità di transizione mancanti, scrivere la matrice di transizione Q associata.
(2) Sapendo che la macchina è inizialmente (al tempo 0) nello stato OFF, qual è la probabilità che al tempo $n = 1$ si trovi in OFF?
(3) Sapendo che la macchina è inizialmente (al tempo 0) nello stato OFF, qual è la probabilità che al tempo $n = 2$ si trovi in OFF?
(4) Il costo c per unità di tempo per mantenere la macchina nello stato OFF è 0, per mantenerla nello stato S è 5 mentre per mantenerla nello stato B è 10. Calcolare il valore atteso e la varianza del costo totale della macchina nei tempi $\{0, 1, 2\}$, sapendo che al tempo 0 si trova nello stato OFF.
(5) Calcolare le distribuzioni invarianti per Q. Calcolare il valore atteso e la varianza del costo c definito sopra nel caso in cui la catena di Markov sia stazionaria.

Una soluzione. 1. Ordiniamo gli stati nel seguente modo: O, S, B e scriviamo la matrice di transizione

$$\begin{pmatrix} ? & 0,2 & 0 \\ ? & 0,4 & 0,4 \\ 0 & ? & 0,4 \end{pmatrix}.$$

Poiché le righe devono sommare a 1, completiamo facilmente ottenendo

$$\begin{pmatrix} 0,8 & 0,2 & 0 \\ 0,2 & 0,4 & 0,4 \\ 0 & 0,6 & 0,4 \end{pmatrix}.$$

2. Basta notare che $P(X_1 = \mathrm{OFF} | \Omega \cap \{X_0 = \mathrm{OFF}\}) = Q_{\mathrm{O} \to \mathrm{O}} = 0,8$.

3. Ricordando i risultati sul calcolo di leggi marginali, basta calcolare il prodotto $(1, 0, 0) \cdot Q^2$ e vederne la componente corrispondente a OFF. Abbiamo

$$(1, 0, 0) \cdot Q = (0.8, 0.2, 0) \cdot Q = (0.68, 0.24, 0.08)$$

da cui $P(X_2 = \mathrm{OFF} | \Omega \cap \{X_0 = \mathrm{OFF}\}) = 0.68$.

4. Consideriamo il vettore colonna $\vec{c} = (0, 5, 10)$. Per calcolare il valore atteso, usiamo i risultati della teoria

$$\begin{aligned}&\mathbb{E}\left[c(X_0) + c(X_1) + c(X_2) | \Omega \cap \{X_0 = \mathrm{OFF}\}\right]\\ &\quad = \mathbb{E}\left[c(X_1) + c(X_2) | \Omega \cap \{X_0 = \mathrm{OFF}\}\right]\\ &\quad = (Q\vec{c})_{\mathrm{OFF}} + \left(Q^2\vec{c}\right)_{\mathrm{OFF}} = 1 + (1,2 + 0,8) = 3.\end{aligned}$$

Per calcolare la varianza, calcoliamo prima il valore atteso del quadrato

$$\begin{aligned}&\mathbb{E}\left[(c(X_1) + c(X_2))^2 | \Omega \cap \{X_0 = \mathrm{OFF}\}\right]\\ &\quad = \mathbb{E}\left[(c(X_1))^2 + (c(X_2))^2 + 2c(X_1)c(X_2) | \Omega \cap \{X_0 = \mathrm{OFF}\}\right].\end{aligned}$$

Spezziamo il valore atteso nella somma dei tre valori attesi e calcoliamo i primi due usando il prodotto matrice per il vettore $\vec{c^2} = (0, 25, 100)$:

$$\mathbb{E}\left[(c(X_1))^2 + (c(X_2))^2 | \Omega \cap \{X_0 = \mathrm{OFF}\}\right] = \left(Q\vec{c^2}\right)_{\mathrm{OFF}} + \left(Q^2\vec{c^2}\right)_{\mathrm{OFF}} = 5 + 14 = 19.$$

Per calcolare il terzo, dobbiamo usare la proprietà di Markov (non abbiamo regole generali per questo tipo di valori attesi). Decomponiamo secondo il sistema di alternative $\{X_1 = i\}$, per $i \in E$:

$$\begin{aligned}&\mathbb{E}\left[c(X_1)c(X_2) | X_0 = \mathrm{O}\right]\\ &= \mathbb{E}\left[c(X_1)c(X_2) | \{X_0 = \mathrm{O}\} \cap \{X_1 = \mathrm{OFF}\}\right] P(X_1 = \mathrm{O} | X_0 = \mathrm{O}) +\\ &+ \mathbb{E}\left[c(X_1)c(X_2) | \{X_0 = \mathrm{O}\} \cap \{X_1 = \mathrm{S}\}\right] P(X_1 = \mathrm{S} | X_0 = \mathrm{O}) +\\ &+ \mathbb{E}\left[c(X_1)c(X_2) | \{X_0 = \mathrm{O}\} \cap \{X_1 = \mathrm{B}\}\right] P(X_1 = \mathrm{B} | X_0 = \mathrm{O})\\ &= \mathbb{E}\left[c(\mathrm{O})c(X_2) | X_1 = \mathrm{O}\right] \cdot 0,8 + \mathbb{E}\left[c(\mathrm{S})c(X_2) | X_1 = \mathrm{S}\right] \cdot 0,2\\ &= 5\mathbb{E}\left[c(X_2) | \{X_1 = \mathrm{S}\}\right] \cdot 0,2 = \mathbb{E}\left[c(X_1) | \{X_0 = \mathrm{S}\}\right]\\ &= (0,4 \cdot 5 + 0,4 \cdot 10) = 6.\end{aligned}$$

In conclusione, il valore atteso del quadrato vale $19 + 2 \cdot 6 = 31$ e la varianza è $31 - 3^2 = 22$, deviazione standard circa $4,7$.

5. Per calcolare le distribuzioni invarianti risolviamo il sistema omogeneo $(Q^\tau - \mathrm{Id})\vec{v} = 0$:

$$Q^\tau - \mathrm{Id} = \begin{pmatrix} -0,2 & 0,2 & 0 \\ 0,2 & -0.6 & 0,6 \\ 0 & 0,4 & -0.6 \end{pmatrix}.$$

Con operazioni di riga otteniamo il sistema equivalente

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 2 & -3 \\ 0 & 0 & 0 \end{pmatrix}.$$

da cui gli autovettori (colonna) di autovalore 1 sono $t(3, 3, 2)$, $(t \neq 0)$. L'unica distribuzione invariante risulta $\vec{\mu} = (3/8, 3/8, 1/4)$. Di conseguenza, possiamo calcolare il valore atteso del costo se la catena è stazionaria:

$$\mathbb{E}\left[c(X_0) | \text{stazionaria}\right] = \frac{3}{8} \cdot 5 + \frac{1}{4} \cdot 10 = 4,375$$

e la varianza calcolando prima il valore atteso del quadrato:

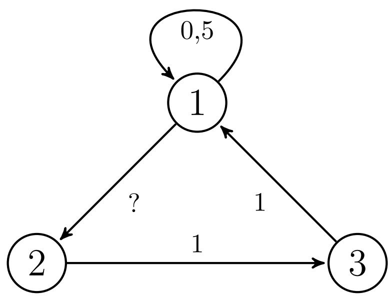
$$\mathbb{E}\left[(c(X_0))^2 | \text{stazionaria}\right] = \frac{3}{8} \cdot 25 + \frac{1}{4} \cdot 100 = 34,375$$

e infine

$$\mathrm{Var}\,(c(X_0)|\text{stazionaria}) = 34,375 - (4.375)^2 = 15,23438$$

$$\sigma(c(X_0)|\text{stazionaria}) \sim 3.90.$$

Problema 31. Consideriamo la catena di Markov rappresentata in figura:



(1) Completare le probabilità di transizione mancanti e scrivere la matrice di transizione Q associata.
(2) Calcolare (aiutandosi eventualmente con un calcolatore) le leggi marginali di X_1, X_2, X_3 ed X_4 sapendo che $X_0 = 1$.
(3) Calcolare le distribuzioni invarianti per Q. Calcolare il valore atteso e la varianza della funzione $f : E \mapsto \mathbb{R}$, $f(e) = e$ nel caso in cui la catena di Markov sia stazionaria.
(4) Nel caso in cui la catena sia stazionaria, calcolare la probabilità che la catena si trovi al tempo 1 nello stato 1, sapendo che al tempo 3 si trova in 1.

Una soluzione. 1. Ordiniamo gli stati nel seguente modo: 1, 2, 3 e scriviamo la matrice di transizione

$$\begin{pmatrix} 0,5 & ? & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Deve essere necessariamente $Q_{1\to 2} = 0,5$.

2. Consideriamo il vettore riga $(1, 0, 0)$, corrispondente alla legge iniziale di X_0, sapendo che $X_0 = 1$. Usiamo risultati noti per ottenere la legge di X_1, X_2, X_3 ed X_4 (sapendo che $X_0 = 1$) in forme di vettori riga, tramite prodotti vettore per matrice

$$(1, 0, 0)Q = (\frac{1}{2}, \frac{1}{2}, 0)$$

$$(1, 0, 0)Q^2 = (\frac{1}{2}, \frac{1}{2}, 0)Q = (\frac{1}{2} \cdot \frac{1}{2}, \frac{1}{2} \cdot \frac{1}{2}, \frac{1}{2}) = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$$

$$(1, 0, 0)Q^3 = (\frac{1}{4}, \frac{1}{4}, \frac{1}{2})Q = (\frac{5}{8}, \frac{1}{8}, \frac{1}{4})$$

$$(1, 0, 0)Q^4 = (\frac{5}{8}, \frac{1}{8}, \frac{1}{4})Q = (\frac{9}{16}, \frac{5}{16}, \frac{1}{8})$$

I risultati si leggono nel seguente modo: ad esempio $(\frac{9}{16}, \frac{5}{16}, \frac{1}{8})$ corrisponde al fatto che

$$P(X_5 = 1|X_0 = 1) = \frac{9}{16}, \quad P(X_5 = 2|X_0 = 1) = \frac{5}{16}, \quad P(X_5 = 2|X_0 = 1) = \frac{1}{8}.$$

3. Per calcolare le distribuzioni invarianti per Q, risolviamo prima il problema di trovare gli autovettori (colonna) di autovalore 1 per Q^τ, ossia risolviamo il sistema omogeneo

$$(Q^\tau - \mathrm{Id})v = 0,$$

cui corrisponde la matrice dei coefficienti

$$\begin{pmatrix} -0,5 & 0 & 1 \\ 0,5 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}$$

Tramite operazioni elementari di riga otttteniamo la matrice dei coefficienti del sistema equivalente

$$\begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

da cui gli autovettori solo della forma $t(2,1,1)$, $t \in \mathbb{R}$ ($t \neq 0$). Come al solito, affinché sia una distribuzione invariante, dobbiamo assicurarci che la somma delle componenti sia 1, da cui $t = \frac{1}{4}$. Perciò abbiamo solamente il caso $\vec{\mu} = \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$.

Calcoliamo il valore atteso di $f(e) = e$, nel caso in cui la catena sia stazionaria. Questo significa che dobbiamo calcolare

$$\begin{aligned}\mathbb{E}\left[f(X_0)|\text{staz.}\right] &= f(1)\vec{\mu}_1 + f(2)\vec{\mu}_2 + f(3)\vec{\mu}_3 \\ &= 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = \frac{7}{4} = 1,75.\end{aligned}$$

Per calcolare la varianza calcoliamo il valore atteso del quadrato:

$$\begin{aligned}\mathbb{E}\left[(f(X_0))^2|\text{staz.}\right] &= (f(1))^2\vec{\mu}_1 + (f(2))^2\vec{\mu}_2 + (f(3))^2\vec{\mu}_3 \\ &= 1^2 \cdot \frac{1}{2} + 2^2 \cdot \frac{1}{4} + 3^2 \cdot \frac{1}{4} = \frac{15}{4} = 3,75.\end{aligned}$$

La varianza vale quindi

$$\mathrm{Var}\left(f(X_0)|\text{staz.}\right) = \frac{15}{4} - \left(\frac{7}{4}\right)^2 = 0,6875.$$

La deviazione standard vale invece $\sqrt{0.6875} \sim 0,83$.

4. Dobbiamo calcolare

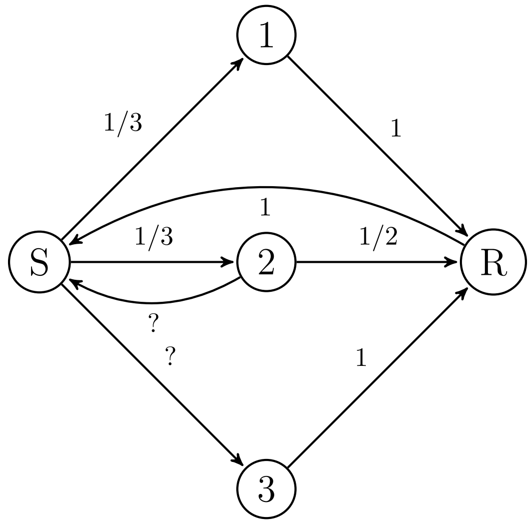
$$\begin{aligned}P\left(X_1 = 1|staz. \cap \{X_3 = 1\}\right) &= P\left(X_3 = 1|X_1 = 1 \cap staz.\right)\frac{P\left(X_1 = 1|staz.\right)}{P\left(X_3 = 1|staz.\right)} \\ &= P\left(X_3 = 1|X_1 = 1 \cap staz.\right)\frac{1/2}{1/2} \\ &= P\left(X_3 = 1|X_1 = 1 \cap staz.\right).\end{aligned}$$

L'informazione (iniziale) che la catena sia stazionaria non ci è di aiuto, perché abbiamo acquisito l'informazione che al tempo 1 la catena si trova in 1, pertanto

$$P\left(X_3 = 1|X_1 = 1 \cap staz.\right) = P\left(X_3 = 1|X_1 = 1\right) = P\left(X_2 = 1|X_0 = 1\right) = \frac{1}{4}.$$

dove abbiamo usato il fatto che la catena di Markov è omogenea (conta solo il numero di transizioni).

Problema 32. Una persona vuole raggiungere da una posizione (S) la posizione R servendosi di uno di tre possibile mezzi, rappresentati in figura da 1, 2 e 3. Il mezzo 2 tuttavia, se scelto, con probabilità 1/2 fallisce e riporta la persona alla situazione in cui deve scegliere. Supponiamo che il comportamento sia modellizzato da una catena di Markov (immaginate voi in quali situazioni potrebbe essere un'ipotesi realistica) che rappresentiamo in figura.



(1) Completare le probabilità di transizione mancanti e scrivere la matrice di transizione Q associata.
(2) Calcolare (aiutandosi eventualmente con un calcolatore) le leggi marginali di X_1, X_2, X_3 ed X_4 sapendo che $X_0 = S$.
(3) Dato $k \in \mathbb{N}$, calcolare la probabilità che la persona, partendo da S al tempo 0, raggiunga lo stato E per la prima volta esattamente al tempo k.
(4) Posto T la variabile aleatoria "primo istante in cui la persona raggiunge E", calcolare il valore atteso di T e la varianza di T sapendo che la persona parte da S al tempo 0.
(5) Calcolare le distribuzioni invarianti di Q.

Una soluzione. 1. Ordiniamo gli stati nel seguente modo: S, 1, 2, 3, R e scriviamo la matrice di transizione (una matrice 5×5):

$$Q = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & ? & 0 \\ 0 & 0 & 0 & 0 & 1 \\ ? & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

da cui necessariamente $Q_{S\to 3} = \frac{1}{3}$, $Q_{2\to S} = \frac{1}{2}$.

2. Come negli esercizi precedenti, consideriamo il vettore riga $\mu = (1, 0, 0, 0, 0)$ corrispondente alla legge al tempo 0 della catena (sapendo che al tempo 0 si trova in S) e calcoliamo prodotti vettore per matrice:

$$\mu Q = (0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0)$$

$$\mu Q^2 = (\mu Q)Q = (\frac{1}{6}, 0, 0, 0, \frac{2}{3} + \frac{1}{6})$$
$$\mu Q^3 = (\frac{5}{6}, \frac{1}{18}, \frac{1}{18}, \frac{1}{18}, 0)$$
$$\mu Q^4 = (\frac{1}{36}, \frac{5}{18}, \frac{5}{18}, \frac{5}{18}, \frac{2}{18} + \frac{1}{36}).$$

Provate a rintracciare a quali cammini corrispondono nel grafo.

3. Poniamo A_k l'evento "la persona raggiunge lo stato R per la prima volta al tempo k". Notiamo che, se vale A_k, allora vale $\{X_k = R\}$ e deve valere $\{X_i \neq R\}$ per ogni $i < k$. Dobbiamo quindi ragionare sulla "geometria" della catene per capire come sono fatti i percorsi possibili che partono da S e raggiungono R in k passi senza mai passare per R (in un tempo precedente a k). Vediamo quindi che le possibilità sono piuttosto ristrette: se il cammino si trova in 1 o 2, deve necessariamente raggiungere R l'istante successivo, quindi gli stati 1 e 2 possono essere "visitati" dalla persona solamente al penultimo istante. Ne segue quindi che la persona deve visitare in sequenza lo stato S ed R alternando così fino al tempo $k-1$, dove deve trovarsi in 1, 2 o 3, quindi al tempo $k-2$ deve trovarsi in S. Ma allora se k è dispari la persona non potrà mai raggiungere in k passi R la prima volta. Abbiamo dunque capito che i cammini possibili che portano in k passi da S ad R senza mai passare per R sono 0 se k è dispari e 3 se k è pari ($k \geq 2$). La probabilità dell'evento è A_k è dunque la somma dei 'pesi' di ciascun cammino da cui

$$P(A_k|X_0 = S) = \begin{cases} 0 & \text{se } k \text{ dispari} \\ \left(\frac{1}{3}\right)^{(k-1)/2} \left(\frac{1}{2}\right)^{(k-1)/2} \cdot \left(\frac{1}{3} + \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2}\right) & \text{se } k \text{ è pari, } k \geq 2. \end{cases}$$

4. La variabile aleatoria è costruita in modo tale che $\{T = k\} = A_k$ dell'esercizio precedente. Perciò la sua legge si trova dal punto precedente. Notiamo allora che T assume valori nei numeri pari $k \geq 2$. Se consideriamo la variabile $T/2$, questa assume valori naturali (positivi) e usando il punto sopra notiamo che la sua legge è geometrica di parametro $p = \frac{5}{6}$ (il parametro di "successo", ossia di scegliere un cammino che porta in due passi da S ad R). Perciò

$$\mathbb{E}\left[T/2|X_0 = S\right] = \frac{1}{p} = \frac{6}{5} \Rightarrow \mathbb{E}\left[T\right] = \frac{12}{5},$$
$$\mathrm{Var}\left(T/2|X_0 = S\right) = \frac{1-p}{p^2} = \frac{6}{25} \Rightarrow \mathrm{Var}\left(T|X_0 = S\right) = \frac{24}{25}.$$

5. Scriviamo la matrice

$$Q^\tau - \mathrm{Id} = \begin{pmatrix} -1 & 0 & \frac{1}{2} & 0 & 1 \\ \frac{1}{3} & -1 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & -1 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & -1 & 0 \\ 0 & 1 & 0,5 & 1 & -1 \end{pmatrix}.$$

Tramite operazioni di riga riduciamo a scalini, ottenendo le matrici

$$\begin{pmatrix} -2 & 0 & 1 & 0 & 2 \\ 1 & -3 & 0 & 0 & 0 \\ 1 & 0 & -3 & 0 & 0 \\ 1 & 0 & 0 & -3 & 0 \\ 0 & 2 & 1 & 2 & -2 \end{pmatrix} \to \begin{pmatrix} 1 & -3 & 0 & 0 & 0 \\ 0 & -6 & 1 & 0 & 2 \\ 0 & 3 & -3 & 0 & 0 \\ 0 & 3 & 0 & -3 & 0 \\ 0 & 2 & 1 & 2 & -2 \end{pmatrix} \to$$

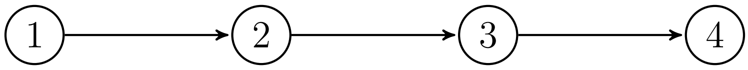
$$\begin{pmatrix} 1 & -3 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & -5 & 0 & 2 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 3 & 2 & -2 \end{pmatrix} \to \begin{pmatrix} 1 & -3 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & -5 & 2 \\ 0 & 0 & 0 & 5 & -2 \end{pmatrix} \to$$

$$\begin{pmatrix} 1 & -3 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 5 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Da cui segue che tutti gli autovettori con autovalore 1 sono della forma $t(6, 2, 2, 2, 5)$, $t \in \mathbb{R}$. Il fatto che la somma delle componenti deve essere uno conduce all'unica distribuzione invariante, data da

$$\vec{\mu} = \frac{1}{17}(6, 2, 2, 2, 5).$$

Problema 33. Un messaggio binario (cioè costituito da una successione di simboli in $\{0, 1\}$) passa attraverso una rete di trasmissione, formata da 4 nodi, rappresentata in figura



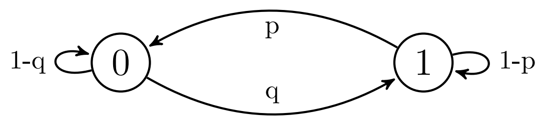
Il messaggio viene spedito dal nodo 1 e viaggia verso il nodo 4: ad ogni passaggio da un nodo all'altro ciascuna cifra viene riportata esattamente oppure trasformata (viene quindi commesso un errore). La struttura della rete è tale per cui, se la cifra nel nodo è 1, la probabilità di errore (ossia venga trasmessa 0) è $p \in (0, 1)$, mentre se la cifra è 0, la probabilità di errore è $q \in (0, 1)$. Supponiamo che il messaggio consista di un solo simbolo, e poniamo $X_i \in \{0, 1\}$ la variabile rappresentante il simbolo ricevuto al nodo $i \in \{1, \ldots, 4\}$.

(1) Il processo X_1, X_2, X_3, X_4 è una catena di Markov omogenea? darne una rappresentazione grafica e scrivere la matrice di transizione.
(2) Sapendo che è stato spedito (dal nodo 1) il simbolo 1, qual è la probabilità che si riceva 1?
(3) Sapendo che è stato ricevuto al nodo 2 il simbolo 0, qual è la probabilità che il nodo 4 riceva 0?
(4) Sapendo che è stato ricevuto al nodo 4 il simbolo 1, qual è la probabilità che sia stato spedito 1? (*Per rispondere a questa domanda potrebbe mancare un'informazione*)

Una soluzione. 1. Sì, il processo è una catena di Markov omogenea, sull'insieme degli stati $\{0, 1\}$.: infatti ciascun nodo riceve l'informazione "presente" e la transmette commettendo o meno un errore, in modo indipendente dal passato. Volendo essere più precisi, la struttura del testo ci lascia intendere che gli errori accadono in modo indipendente in ciascun nodo (eccetto il fatto che la probabilità di errore dipende dallo stato del messaggio). Poiché lo stato futuro del messaggio messaggio è funzione dell'errore e dello stato presente (a sua volta funzione del messaggio iniziale e della sequenza di possibili errori), conoscere tutta la storia passata non influenza la nostra previsione rispetto alla nostra conoscenza del solo stato presente. Le probabilità di errore inoltre sono indipendenti dal tempo, quindi il processo è omogeneo. Scriviamo le probabilità di transizione come matrice, ordinando gli stati nell'ordine 0, 1:

$$Q = \begin{pmatrix} 1-q & q \\ p & 1-p \end{pmatrix}.$$

In forma grafica,



È richiesta la probabilità $P(X_4 = 1|X_1 = 1)$ Calcoliamola con l'aiuto del prodotto vettore matrice, ponendo $\vec{\mu} = (0,1)$ e quindi calcoliamo $\vec{\mu}Q^3$:

$$(0,1)Q = (p, 1-p), \quad (0,1)Q^2 = (p, 1-p)Q = (p(1-q)+p(1-p), pq+(1-p)^2)$$

$$\begin{aligned}(0,1)Q^3 =&(p(1-q)+p(1-p), pq+(1-p)^2)Q \\ &= ((p(1-q)+p(1-p))(1-q) + \left(pq+(1-p)^2\right)p, \\ &\quad (p(1-q)+p(1-p))\,q + \left(pq+(1-p)^2\right)(1-p)).\end{aligned}$$

La risposta è dunque

$$P(X_4 = 1|X_1 = 1) = (p(1-q)+p(1-p))\,q + \left(pq+(1-p)^2\right)(1-p).$$

3. In questo caso sappiamo che $X_2 = 0$, quindi ci basta calcolare $P(X_4 = 0|X_2 = 0)$, stavolta ponendo $\vec{\mu} = (1,0)$:

$$(1,0)Q = (1-q, q), \quad (0,1)Q^2 = (1-q, q)Q = ((1-q)^2 + qp, \dots)$$

da cui $P(X_4 = 0|X_2 = 0) = (1-q)^2 + qp$.

4. Usiamo la formula di Bayes

$$P(X_1 = 1|X_4 = 1) = P(X_4 = 1|X_1 = 1)\frac{P(X_1 = 1|\Omega)}{P(X_4 = 1|\Omega)}.$$

Il problema è che mentre conosciamo $P(X_4 = 0|X_1 = 0)$ (calcolata prima) non conosciamo le probabilità marginali rispetto all'informazione iniziale. A questo punto dobbiamo chiederci: qual è la legge di X_0 rispetto all'informazione iniziale? (perché di conseguenza la legge di X_4 si calcola). Abbiamo due possibili risposte: in un primo momento, visto che gli stati sono due, possiamo usare il principio di indifferenza e porre $P(X_0 = 0|\Omega) = 1/2 = P(X_1 = 0|\Omega)$. Una seconda risposta è di calcolare le distribuzioni invarianti e supporre quindi che la catena sia stazionaria: questa seconda risposta ci permette di tenere conto che gli stati 0 ed 1 non sono necessariamente equivalenti (infatti la probabilità stazionaria non sarà uniforme se $p \neq q$). Se scegliamo questa seconda strada la risposta allora è molto semplice, perché il quoziente si semplifica ad 1, e otteniamo

$$P(X_1 = 1|X_4 = 1) = (p(1-q)+p(1-p))\,q + \left(pq+(1-p)^2\right)(1-p).$$

7. Richiami di algebra lineare

Per risolvere i problemi relativi alle catene di Markov abbiamo visto che un po' di algebra lineare può essere utile. A tale scopo richiamiamo qui le definizioni principali, e l'algoritmo di riduzione di Gauss per risolvere sistemi lineari. Rimandiamo a qualunque testo per maggiori dettagli: inoltre non diamo le definizioni nella massima generalità.

Vettori e matrici. Chiamiamo vettore colonna $\vec{v} \in \mathbb{R}^k$ una collezione ordinata di k numeri reali (disposti a colonna)

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{k-1} \\ v_k \end{pmatrix}.$$

Un vettore riga $\vec{\mu} \in \mathbb{R}^k$ è sempre una collezione ordinata di k numeri reali, ma disposti a riga:

$$\vec{\mu} = \begin{pmatrix} \mu_1 & \mu_2 & \dots & \mu_{k-1} & \mu_k \end{pmatrix}.$$

In effetti la distinzione matematica tra i due tipi di vettore è meglio messa in evidenza introducendo il concetto più generale di matrice a r ($r \geq 1$) righe e c ($c \geq 1$) colonne, $M \in \mathbb{R}^{r \times c}$, che è definita come una collezione ordinata di $r \times c$ numeri reali $(M_{i,j})_{i=1,\dots,r}^{j=1,\dots,c}$:

$$M = \begin{pmatrix} M_{1,1} & M_{1,2} & \dots & M_{1,c-1} & M_{1,c} \\ M_{2,1} & M_{1,2} & \dots & M_{1,c-1} & M_{1,c} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ M_{r,1} & M_{r,2} & \dots & M_{r,c-1} & M_{r,c} \end{pmatrix}$$

Quindi la componente $M_{i,j}$ indica il numero reale nella posizione della i-esima riga e j-esima colonna (notate che è invertito rispetto alle coordinate cartesiane, in cui (x, y) indica prima la componente orizzontale x e poi la componente verticale y). Nel caso di matrici di transizione abbiamo usato la notazione $M_{i \to j}$, più espressiva.

Possiamo quindi identificare i vettori riga con matrici con $r = 1$, e i vettori colonna con matrici con $c = 1$. Questa identificazione ci permette di sfruttare il prodotto generale *righe per colonne* tra due matrici, $M \in \mathbb{R}^{r \times b}$, $N \in \mathbb{R}^{b \times c}$, definito mediante la formula

$$(M \cdot N)_{i,j} = \sum_{k=1}^{b} M_{i,k} N_{k,j}, \quad \text{per } i \in \{1, \dots, r\},\ j \in \{1, \dots, c\}.$$

Proviamo graficamente ad evidenziare l'operazione descritta dalla formula sopra:

$$\begin{pmatrix} M_{1,1} & M_{1,2} & \dots & M_{1,b-1} & M_{1,b} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \color{red}{M_{i,1}} & \color{red}{M_{i,2}} & \color{red}{\dots} & \color{red}{M_{i,b-1}} & \color{red}{M_{i,b}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ M_{r,1} & M_{r,2} & \dots & M_{r,b-1} & M_{r,b} \end{pmatrix} \begin{pmatrix} N_{1,1} & \dots & \color{blue}{N_{1,j}} & \dots & N_{1,c} \\ \vdots & \ddots & \color{blue}{\vdots} & \ddots & \vdots \\ N_{2,1} & \dots & \color{blue}{N_{2,j}} & \dots & N_{2,c} \\ \vdots & \ddots & \color{blue}{\vdots} & \ddots & \vdots \\ N_{b,1} & \dots & \color{blue}{N_{b,j}} & \dots & N_{b,c} \end{pmatrix}$$

dobbiamo moltiplicare ciascun termine $\color{red}{M_{i,k}}$ con il corrispondente $\color{blue}{M_{k,j}}$, e sommare tutti i prodotti ottenendo $(M \cdot N)_{i,j} = \sum_{k=1}^{b} \color{red}{M_{i,k}}\color{blue}{N_{k,j}}$.

Notate che il numero di colonne di M deve coincidere con quello delle righe di N, altrimenti non ri riesce proprio a definire un prodotto con la formula sopra! Inoltre il prodotto è associativo (quando ha senso fare il prodotto di tre o più matrici).

Dato $k \geq 1$ matrice identica $\mathrm{Id} \in \mathbb{R}^{k \times k}$ è definita come la matrice con tutte le entrate 0 eccetto sulla diagonale (ossia i termini con $i = j$), in cui vale 1:

$$\mathrm{Id} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

Data una matrice $M \in \mathbb{R}^{r \times c}$, la sua *trasposta* è definita come la matrice $M^\tau \in \mathbb{R}^{c \times r}$, $M^\tau_{i,j} := M_{j,i}$. L'effetto grafico è che operiamo una simmetria rispetto alla diagonale (ossia i termini con $i = j$).

Sistemi lineari. C'è una importante corrispondenza tra matrici, vettori e operazioni tra questi e la risoluzione dei sistemi lineari, ossia il problema di risolvere simultaneamente un certo numero di equazioni in cui le incognite (anche più di una) appaiono come polinomi di grado al più 1. Infatti un sistema lineare di $r \geq 1$ equazioni in k *incognite* $x_1, x_2, \dots, x_k$ si presenta (o può essere ricondotto ad un sistema equivalente, "portando a destra o a sinistra" dell'uguaglianza i termini opportuni) nell forma

$$\begin{cases} a_{1,1}x_1 + & a_{1,2}x_2 + & \dots + & a_{1,k-1}x_{k-1} & a_{1,k}x_k & = b_1 & \text{(prima equazione)} \\ a_{2,1}x_1 + & a_{2,2}x_2 + & \dots + & a_{2,k-1}x_{k-1} & a_{2,k}x_k & = b_1 & \text{(seconda equazione)} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \\ a_{r-1,1}x_1 + & a_{r-1,2}x_2 + & \dots + & a_{r-1,k-1}x_{k-1} & a_{r-1,k}x_k & = b_{r-1} & (r-1\text{-esima equazione}) \\ a_{r,1}x_1 + & a_{r,2}x_2 + & \dots + & a_{r,k-1}x_{k-1} & a_{r,k}x_k & = b_r & (r\text{-esima equazione}) \end{cases}$$

Possiamo trovare una notazione più compatta tenendo conto delle incognite in modo "implicito" (dato dalla posizione) raccogliendo i *coefficienti* $a_{i,j}$ e i *termini noti* b_i in una matrice con r righe (tante quante le equazioni) e $k+1$ colonne (una per ciascuna incognita e l'ultima per i termini noti) definita così:

$$\left(\begin{array}{ccccc|c} a_{1,1} & a_{1,2} & \dots & a_{1,k-1} & a_{1,k} & b_1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,k-1} & a_{2,k} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a_{r-1,1} & a_{r-1,2} & \dots & a_{r-1,k-1} & a_{r-1,k} & b_{r-1} \\ a_{r,1} & a_{r,2} & \dots & a_{r,k-1} & a_{r,k} & b_r \end{array}\right)$$

La barra verticale serve a separare (graficamente) i termini corrispondenti ai membri a destra. Quando $b_1 = b_2 = \ldots = b_k = 0$, il sistema è detto *omogeneo* e non si scrive solitamente la colonna di zeri a destra (sarebbe inchiostro sprecato). Data una matrice $A \in \mathbb{R}^{r\times k}$ e un vettore colonna $b \in \mathbb{R}^k$, possiamo considerare il sistema corrispondente alla matrice

$$(A|b)\,,$$

che ha r equazioni in k incognite. Possiamo anche formulare tale sistema usando il prodotto di matrice: trovare una soluzione per il sistema significa trovare un vettore colonna $x = (x_1, \ldots, x_k) \in \mathbb{R}^k$ tale che valga l'identità

$$A \cdot x = b.$$

Provate a rendervene conto: basta osservare che i membri a sinistra delle equazioni del sistema scritto sopra sono proprio le righe risultanti dal prodotto righe per colonne di $A \cdot x$.

Questa corrispondenza tra matrici e sistemi permette agevolmente di sfruttare l'algoritmo di eliminazione gaussiana, una tecnica risolutiva efficace per un sistema lineare. Questa si basa su due fatti: il primo è che un sistema è equivalente ad un secondo sistema (ossia hanno esattamente le stesse soluzioni) se il secondo è ottenuto dal primo se trasformiamo un'equazione aggiungendo (membro a membro) un multiplo di un'altra equazione presente nel sistema. Il secondo fatto è che è molto semplice risolvere sistemi in forma *triangolare*, ossia in cui la matrice associata è della forma *a gradini*, cioè del tipo

$$\left(\begin{array}{ccccc|c} a_{1,1} & a_{1,2} & \dots & a_{1,k-1} & a_{1,k} & b_1 \\ 0 & a_{2,2} & \dots & a_{2,k-1} & a_{2,k} & b_2 \\ 0 & 0 & \dots & a_{3,k-1} & a_{3,k} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & a_{r-1,k-1} & a_{r-1,k} & b_{r-1} \\ 0 & 0 & \dots & 0 & a_{r,k} & b_r \end{array}\right)$$

infatti è possibile ricavare in sequenza tutte le soluzioni, partendo dalle variabili "libere" (che e "risalendo" fino ad otterele tutte. In realtà la forma a gradini più generale potrebbe avere delle righe di zeri, che però non danno alcun problema (corrispondono all'equazione banale $0 = 0$), oppure delle righe corrispondenti all'equazione $0 = b_i$, che è impossibile se $b_i = 0$; in caso di dubbi ricordate sempre la corrispondenza tra la matrice e il sistema che volete risolvere.

L'idea di Gauss è di procedere sistematicamente tramite queste operazioni per ottenere una forma a gradini. Notiamo che altre operazioni ammesse sono: scambiare due righe tra loro, e di moltiplicare una riga (ossia l'equazione ambo i membri) per un numero diverso da zero.

Non diamo la definizione precisa dell'algoritmo, ma ragioniamo su un esempio. Consideriamo la matrice

$$\begin{pmatrix} -0,5 & 0 & 1 \\ 0,5 & -1 & 0 \\ 0 & 2 & -2 \end{pmatrix}$$

che corrisponde a un sistema omogeneo di tre equazioni in tre incognite. Volendo "eliminare" il più possibile le componenti della prima colonna, basta sommare alla seconda riga la prima riga, ottenendo la matrice (corrispondente a un sistema equivalente)

$$\begin{pmatrix} -0,5 & 0 & 1 \\ 0,5-0,5 & -1+0 & 0+1 \\ 0 & 2 & -2 \end{pmatrix} = \begin{pmatrix} -0,5 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 2 & -2 \end{pmatrix}$$

a questo punto la prima colonna è sistemata. Passiamo alla seconda colonna: volendo eliminare il più possibile (per ottenere una forma a gradini), possiamo sommare alla terza riga due volte la seconda riga, ottenendo

$$\begin{pmatrix} -0,5 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 2+2\cdot(-1) & -2+2\cdot 1 \end{pmatrix} = \begin{pmatrix} -0,5 & 0 & 1 \\ 0 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

vedete che l'operazione non distrugge la strutta parziale a gradini ottenuta nella prima colonna (questo è fondamentale!). A questo punto siamo già in forma a gradini, e le equazioni sono due (la terza è $0 = 0$), precisamente $-0,5x + z = 0$, $-y + z = 0$. Per risolverle, possiamo scegliere $z = t$ (come preferiamo) e ottenere (dal basso verso l'alto) $y = t$, $x = 2t$.

Autovettori. Diamo la seguente definizione: data una matrice $A \in \mathbb{R}^{k\times k}$ (quadtrata!) un vettore colonna $v \in \mathbb{R}^k \setminus \{0\}$ è detto autovettore per Q se esiste $\lambda \in \mathbb{R}$ tale che

$$A \cdot v = \lambda v,$$

dove $(\lambda v)_i = \lambda v_i$, per $i \in \{1, \ldots, k\}$. Il numero λ è detto *autovalore* relativo a v. Notiamo che, se Id indica la matrice identica, possiamo scrivere la relazione sopra nel seguente modo:

$$0 = A \cdot v - \lambda \operatorname{Id} \cdot v = (A - \lambda \operatorname{Id}) \cdot v,$$

ossia v (le sue componenti) risolvono il sistema lineare omogeneo con matrice dei coefficienti $A - \lambda \operatorname{Id}$. Questo ci permette di calcolare autovettori per cui un certo $\lambda \in \mathbb{R}$ è il loro autovalore. Nel nostro caso (calcolo di distribuzioni invarianti), vale $\lambda = 1$, $A = Q^\tau$ e quindi ci riconduciamo al sistema lineare omogeneo $(Q^\tau - \operatorname{Id})x = 0$.

Università degli studi di Pisa. E-mail: dario.trevisan@unipi.it