Comportamento asintotico delle Catene di Markov

In queste note analizzeremo il comportamento asintotico della catene di Markov a tempo discreto omogenee, con spazio degli stati di dimensione *finita.* I risultati che deriviamo hanno una dimostrazione alternativa utilizzando i risultati noti relativi a matrici ad elementi positivi (in particolare, il Teorema di Perron-Frobenius). La derivazione qui presentata utilizza un semplice approccio alternativo, che non utilizza tale teoria, suggerito dall'impostazione in [1].

Consideremo il caso di CdM *irriducibili*, ovvero tali che, presi comunque due stati i, j, sia non nulla la probabilità di raggiungere lo stato j, partendo dallo stato i, in un numero finito di passi. Questo equivale al fatto che, nel grafo che rappresenta il diagramma di transizione della CdM, esista un percorso $i \rightarrow j$ che porta dallo stato i allo stato j.

Nella prima parte di queste note, consideremo il caso di CdM con matrice di transizione ad un passo con elementi *strettamente* positivi, e faremo vedere che in queste ipotesi, si ha, per ogni vettore delle probabilità iniziali $\mathbf{p}(0)$,

$$\lim_{n\to+\infty} \mathbf{p}(0)\mathbf{P}^n = \mathbf{p},$$

con $\mathbf{p}$ distribuzione stazionaria della catena di Markov.

Faremo poi vedere che i risultati si estendono al caso delle CdM irriducibili e aperiodiche in cui $\mathbf{P}$ non ha necessariamente elementi strettamente maggiori di zero.

Il caso di P ad elementi strettamente maggiori di zero

Il risultato si basa sui lemmi seguenti.

Lemma 1 *Sia* $\mathbf{c} = [c_1, ..., c_M]^T$ *un vettore colonna, con valore massimo delle componenti pari a* M *e valore minimo pari a* m. *Sia* $\mathbf{p} = [p_1, ..., p_M]$ *un vettore riga, con componenti* $0 \le d \le p_i \le 1$, $\sum_i p_i = 1$. *Allora*

$$\mathbf{pc} = \sum_{i=1}^{M} p_i c_i \le dm + (1-d)M, \tag{1}$$

$$\mathbf{pc} = \sum_{i=1}^{M} p_i c_i \ge dM + (1-d)m. \tag{2}$$

Prova Siano h e k gli indici (non necessariamente unici), delle componenti di $\mathbf{c}$ che corrispondono al valore massimo e minimo, $c_h = m$, $c_k = M$. Per dimostrare la (1), possiamo scrivere

$$\sum_i p_i c_i \quad = \quad p_h m + \sum_{i \neq h} p_i c_i$$

$$
\begin{aligned}
&\leq \quad p_h m + M \sum_{i \neq h} p_i = (p_h - d)m + (1 - p_h)M + dm \\
&\leq \quad (p_h - d)M + (1 - p_h)M + dm = dm + (1 - d)M,
\end{aligned}
$$

dove l'ultima disuguaglianza segue dal fatto che, essendo $p_h - d \geq 0$, si ha $(p_h - d)m \leq (p_h - d)M$.

Per la (2), possiamo scrivere

$$
\begin{aligned}
\sum_i p_i c_i &= p_k M + \sum_{i \neq k} p_i c_i \\
&\geq p_k M + m \sum_{i \neq k} p_i = (p_k - d)M + (1 - p_k)m + dM \\
&\geq (p_k - d)m + (1 - p_k)m + dM = dM + (1 - d)m.
\end{aligned}
$$

□

Dal lemma precedente segue immediatamente il risultato seguente.

Lemma 2 *Sia* $\mathbf{P}$ *una matrice stocastica le cui componenti verificano* $0 \leq d \leq P_{i,j} \leq 1$. *Sia* $\mathbf{c}$ *un vettore colonna, con valore massimo delle componenti pari a* M_c *e valore minimo pari a* m_c. *Allora il vettore* $\mathbf{y} = \mathbf{Pc}$ *ha componenti con valore massimo* $M_y \leq M_c$ *e minimo* $m_y \geq m_c$. *Inoltre* $M_y - m_y \leq (1 - 2d)(M_c - m_c)$.

Prova Dal risultato del Lemma 1, si deduce che

$$
M_y \leq dm_c + (1 - d)M_c,
$$

$$
m_y \geq dM_c + (1 - d)m_c.
$$

Il risultato del Lemma 1 vale in particolare per $d = 0$, da cui segue $M_y \leq M_c$, $m_y \geq m_c$. Il risultato relativo alla differenza segue per sottrazione delle due relazioni precedenti. □

Sia dunque $\mathbf{P}$ la matrice di transizione di una catena di Markov con elementi strettamente positivi. Tale catena è ovviamente irriducibile. Inoltre, i risultati precedenti permettono di concludere che $\mathbf{P}^n$ ha tutti elementi strettamente positivi per ogni n, per cui la catena è anche aperiodica (si tratta dunque di un caso particolare di CdM irriducibile e aperiodica). Detto d il valore minimo delle componenti di $\mathbf{P}$, si noti che deve essere $0 < d \leq 1/2$ (altrimenti almeno una riga di $\mathbf{P}$ sommerebbe ad un valore maggiore di 1). Il Lemma 2 dimostra che la differenza fra il valore massimo e minimo delle componenti di ciascuna colonna di $\mathbf{P}^{n+1}$ descresce con il fattore $(1 - 2d)^n$. In particolare, la differenza non può crescere, e anzi, se d è *strettamente* maggiore di zero, tale differenza tende a 0, in quanto $(1 - 2d) < 1$. Questo, unito al fatto che il valore massimo di ciascuna colonna di $\mathbf{P}^{n+1}$ non può crescere, e che il valore minimo non può diminuire al crescere di n, permette di concludere che gli elementi di ciascuna colonna di $\mathbf{P}^{n+1}$ tendono ad un valore costante e maggiore di zero per $n \to +\infty$. Ne segue il seguente teorema.

Teorema 1 *Sia* $\mathbf{P}$ *la matrice di transizione ad un passo di una CdM finita e omogenea, con elementi tutti strettamente positivi (la catena è un caso particolare di CdM irriducibile e aperiodica). Allora*

$$\lim_{n\to+\infty} \mathbf{P}^n = \begin{bmatrix} \mathbf{p} \\ \vdots \\ \mathbf{p} \end{bmatrix}.$$

Ne consegue che, per ogni vettore delle probabilità iniziali $\mathbf{p}(0)$*, si ha*

$$\lim_{n\to+\infty} \mathbf{p}(0)\mathbf{P}^n = \mathbf{p},$$

con $\mathbf{p}$ *distribuzione stazionaria della catena di Markov.*

Il caso generale delle CdM irriducibili e aperiodiche

Si ricorda che il periodo d_i dello stato i di una CdM, è definito come il massimo comun divisore delle lunghezze dei percorsi $i \to i$ (MCD$_i$).

In vista della dimostrazione del successivo teorema, dimostriamo il seguente lemma.

Lemma 3 *Supponiamo che* D *divida* h *e* k*, con* $k \geq h$*. Allora* D *divide anche* $k - h$*.*

Prova Si ha per ipotesi $h = mD$ e $k = nD$, $n \geq m$. Dunque $k - h = (n - m)D$. □

Teorema 2 *In una catena di Markov irriducibile, tutti gli stati hanno lo stesso periodo* d*. In altre parole, se le lunghezze dei percorsi* $i \to i$ *hanno massimo comun divisore MCD*$_i$*, allora, preso uno stato qualsiasi* j *diverso da* i*, anche le lunghezze dei percorsi* $j \to j$ *hanno massimo comun divisore MCD*$_i$*=MCD*$_j$*.*

Prova Dato che la catena è irriducibile, esiste un percorso $i \to j$ di lunghezza h e un percorso $j \to i$ di lunghezza k. Si consideri un qualsiasi percorso $j \to j$ (ne esiste almeno uno, dato che la catena è irriducibile), e sia l la sua lunghezza. Ne segue che esistono due percorsi $i \to i$ di lunghezza $h+k+l$ e $h+k$, rispettivamente, entrambi divisibili, per ipotesi, per MCD$_i$. In base al lemma precedente, risulta che l è anch'esso divisibile per MCD$_i$. Dunque, ogni percorso $j \to j$ ha una lunghezza divisibile per MCD$_i$, cosicché MCD$_j \geq$ MCD$_i$. Scambiando il ruolo di i e j si ottiene MCD$_i \geq$ MCD$_j$, da cui MCD$_j$ = MCD$_i$. □

Si consideri ora una catena di Markov irriducibile e aperiodica. Riferendoci al diagramma di transizione, notiamo che se esiste un percorso di lunghezza n da uno stato i allo stato j, allora $\mathrm{P}[x_n = j | x_0 = i]$ è strettamente maggiore di zero. Facciamo ora vedere che, nel caso di una CdM aperiodica, *per ogni* n, purché sufficientemente grande, esiste un percorso $i \to i$ di lunghezza n (e dunque anche un percorso $i \to j$, per i, j arbitrari, tramite un percorso $i \to i$ e poi $i \to j$).

Prova Siano p e q le lunghezze di due percorsi $i \to i$. Dato che il periodo della catena di Markov è 1, possiamo scegliere p e q in modo che non abbiano divisori maggiori di 1 in comune. L'algoritmo di Euclide per il calcolo del massimo comun divisore, permette di determinare h'_0 e k'_0 (necessariamente di segno discorde), tali che

$$h'_0 p + k'_0 q = 1.$$

Di conseguenza, per ogni n, possiamo determinare una soluzione dell'equazione

$$h_0 p + k_0 q = n \tag{3}$$

ponendo $h_0 = nh'_0$, $k_0 = nk'_0$, con h_0 e k_0 di segno discorde. Occorre dimostrare che, per n sufficientemente grande, esiste invece una soluzione di

$$hp + kq = n, \tag{4}$$

con $h \geq 0$, $k \geq 0$: infatti, potremmo determinare in questo caso un percorso $i \to i$ di lunghezza n, percorrendo h volte il percorso $i \to i$ di lunghezza p e k volte il percorso $i \to i$ di lunghezza q.

Al variare di h e k, la (4) definisce un insieme discreto di punti sulla retta di equazione $xp + yq = n$. Partendo dalla soluzione (3), i punti dell'insieme si ottengono sottraendo a k_0 il valore mp, multiplo di p, e aggiungendo mq a h_0, per ogni m intero, ottenendo infatti (vedi Fig. 1).

$$(h_0 + mq)p + (k_0 - mp)q = n.$$

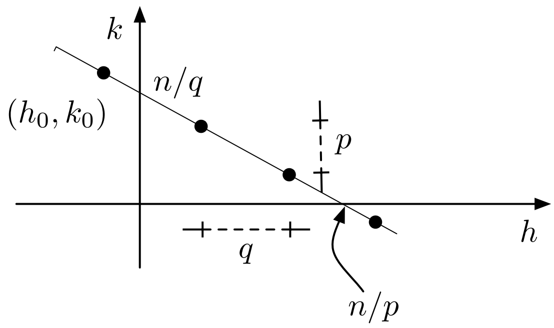

Figura 1: Le soluzioni dell'equazione (4).

Possiamo dunque soddisfare l'Eq. (4) con h e k maggiori o uguali a zero se almeno uno dei punti dell'insieme discreto si trova nel primo quadrante. Questo è garantito se la distanza fra due punti consecutivi sulla retta è minore o uguale della lunghezza del segmento di retta contenuto nel primo quadrante, ovvero se

$$p^2 + q^2 \leq \left(\frac{n}{q}\right)^2 + \left(\frac{n}{p}\right)^2,$$

da cui si ricava

$$n \geq pq.$$

Si tratta ovviamente di un vincolo sufficiente su n. $\square$

Il risultato precedente ci permette di concludere che, per una catena di Markov irriducibile e aperiodica di dimensione finita, esiste un N sufficientemente grande tale da rendere strettamente positiva la probabilità di raggiungere lo stato j, partendo dallo stato i, per ogni coppia di stati i, j. In altre parole, possiamo trovare N tale che la matrice di transizione a N passi $\mathbf{Q} = \mathbf{P}^N$ ha elementi *strettamente* positivi. I risultati della sezione precedente, relativi a matrici di transizione a elementi positivi, ci permettono di concludere che $\mathbf{Q}^n$ tende ad una matrice con colonne costanti per $n \to +\infty$. Questo risultato, unito al fatto che in generale la differenza fra il minimo e il massimo valore delle componenti in ciascuna colonna in $\mathbf{P}^n$ non può crescere, permette di dedurre il seguente teorema.

Teorema 3 *Si consideri una CdM omogenea, di dimensione finita, irriducibile e aperiodica. Allora*

$$\lim_{n\to+\infty} \mathbf{P}^n = \begin{bmatrix} \mathbf{p} \\ \vdots \\ \mathbf{p} \end{bmatrix}.$$

Ne consegue che, per ogni vettore delle probabilità iniziali $\mathbf{p}(0)$*, si ha*

$$\lim_{n\to+\infty} \mathbf{p}(0)\mathbf{P}^n = \mathbf{p},$$

con $\mathbf{p}$ *distribuzione stazionaria della catena di Markov.*

Riferimenti bibliografici

[1] “Grinstead and Snells Introduction to Probability: The CHANCE Project,” `http://www.dartmouth.edu/~chance/`