

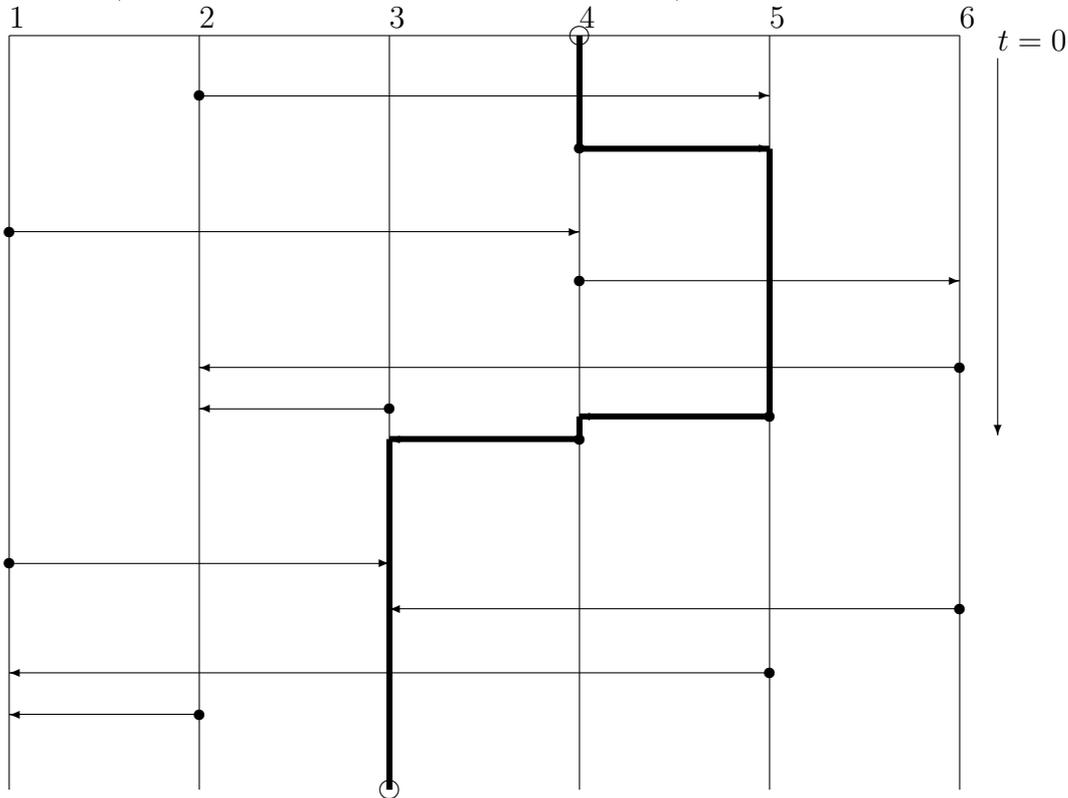
CATENE DI MARKOV A TEMPO CONTINUO

1 Costruzione Grafica

Sia dato A un insieme finito. Per ogni elemento $a \in A$ consideriamo un processo di Poisson $N^a(t)$ di parametro λ_a . I processi di Poisson sono indipendenti fra loro. Per ogni elemento $a \in A$ consideriamo anche una sequenza i.i.d. di variabili casuali Y_i^a con $i \in \mathbb{N}$. Le variabili casuali Y_i^a assumono valori in A ed hanno distribuzione

$$\mathbb{P}(Y_i^a = b) = p_a(b), \quad b \in A. \quad (1.1)$$

Per ogni $a \in A$ abbiamo che $p_a(\cdot)$ e' una misura di probabilit  su A , ossia $p_a(b) \geq 0$ ed inoltre $\sum_{b \in A} p_a(b) = 1$. Le variabili casuali associate a due elementi diversi di A sono tra di loro indipendenti. Tutte le variabili casuali sono inoltre indipendenti dai processi di Poisson sopra menzionati. Costruiamo ora a partire da questi elementi un processo stocastico $X(t)$ avente come spazio delle traiettorie $D([0, T], A)$, lo spazio delle funzioni CADLAG (continue a destra aventi limite a sinistra) a valori in A .



La figura sopra illustra un esempio di costruzione grafica che ora spiegheremo. Si tratta di un caso con $A = \{1, 2, 3, 4, 5, 6\}$. Per ogni elemento di A abbiamo disegnato una retta verticale che rappresenta l'asse dei tempi orientato verso il basso. In ciascuno di questi assi temporali abbiamo rappresentato gli eventi del corrispondente processo di Poisson con dei pallini \bullet . Le frecce disegnate corrispondono invece ai valori delle variabili casuali Y . In corrispondenza del k -esimo evento (rispetto all'ordine crescente dei tempi) del processo di Poisson associato allo stato $a \in A$ disegniamo una freccia che esce da a

ed entra in $b \in A$ se la variabile casuale Y_k^a assume il valore b (si noti che in principio la variabile casuale Y_k^a potrebbe assumere il valore a ; in tal caso non si disegna nessuna freccia.). Scegliamo ora una condizione iniziale $a \in A$ e costruiamo il processo associato a questa condizione iniziale, che verrà indicato con $X^a(t)$. Nel caso specifico dell'esempio in figura abbiamo scelto come condizione iniziale l'elemento 4 e tale scelta è rappresentata dal pallino \circ disegnato all'istante $t = 0$ sulla retta dei tempi associata allo stato 4. Il significato della costruzione grafica è ora molto semplice: al crescere del tempo t il pallino \circ evolve muovendosi lungo le rette verticali e saltando da uno stato all'altro tutte le volte che incontra una freccia orientata dallo stato in cui si trova il pallino verso un'altro stato. Abbiamo che $X^a(t) = b$ se il pallino \circ si trova nello stato b al tempo t .

È facile convincersi del fatto che tale evoluzione è ben definita se non esistono due eventi \bullet che avvengono nello stesso istante temporale (in questo caso ad esempio potremmo avere in un istante t una freccia uscente dallo stato a ed entrante nello stato b ed una freccia uscente dallo stato b ed entrante nello stato a e l'evoluzione di \circ non sarebbe definita).

Poiché gli eventi \bullet sono associati a processi di Poisson indipendenti la situazione patologica sopra illustrata avviene con probabilità zero. Infatti dalla teoria del processo di Poisson sappiamo che il processo $N(t) := \sum_{a \in A} N^a(t)$, è un processo di Poisson di parametro $\lambda = \sum_{a \in A} \lambda_a$. Il verificarsi di due eventi \bullet nello stesso istante temporale comporta che il processo N abbia un salto di due unità, e questo sappiamo essere un evento di probabilità zero per un processo di Poisson.

Una volta definito il processo X^a associato ad una fissata condizione iniziale possiamo definire in modo naturale un processo X^μ associato alla condizione iniziale μ misura di probabilità su A . Questo lo si ottiene scegliendo il dato iniziale come $X^\mu(0) = I$ dove I è una variabile casuale a valori in A avente distribuzione μ ed indipendente da tutte le variabili casuali utilizzate nella costruzione grafica. Una volta scelta la condizione iniziale in questo modo l'evoluzione successiva avviene sempre secondo le regole sopra descritte. Dati t_1, \dots, t_n dei tempi ed a_1, \dots, a_n degli elementi di A otteniamo quindi

$$\begin{aligned} \mathbb{P}(X^\mu(t_1) = a_1, \dots, X^\mu(t_n) = a_n) &= \sum_{a \in A} \mathbb{P}(X^\mu(t_1) = a_1, \dots, X^\mu(t_n) = a_n | I = a) \mathbb{P}(I = a) \\ &= \sum_{a \in A} \mu(a) \mathbb{P}(X^a(t_1) = a_1, \dots, X^a(t_n) = a_n). \end{aligned} \quad (1.2)$$

In questo calcolo abbiamo usato l'indipendenza della variabile casuale I dalle variabili casuali utilizzate per la costruzione grafica, il fatto che quando $I = a$ l'evoluzione di X^μ coincide con quella di X^a e la distribuzione di I . Un processo così costruito tramite questa costruzione grafica si dice una *Catena di Markov a tempo continuo*. In particolare il processo appena descritto è una catena di Markov a tempo continuo omogenea temporalmente. Non discuteremo qui il caso non omogeneo temporalmente.

Possiamo ora immaginare di costruire il diagramma con punti e frecce della costruzione grafica ma di far partire il pallino \circ invece che dal tempo $t = 0$ da un istante fissato successivo T . Questo pallino evolverà sempre secondo le regole della costruzione grafica e determinerà un processo stocastico a valori in A (che indichiamo con X_T^a se si pone il pallino all'istante T nella posizione a) il cui stato al tempo $t \geq 0$ sarà determinato dalla posizione del pallino al tempo $T + t$ rispetto al riferimento temporale della costruzione

grafica. Lungo l'asse temporale associato allo stato $a \in A$ l'evoluzione e' determinata dal processo $N^a(t+T) - N^a(T)$ che, in base all'esercizio 1 dei complementi sul processo di Poisson, e' un processo di Poisson di parametro λ_a indipendente da $[N^a(s)]_{s \in [0, T]}$. Le configurazioni delle frecce uscenti dallo stato a sono inoltre determinate dalle variabili casuali $Z_i^a := Y_{N^a(T)+i}^a$ che sono i.i.d. con distribuzione $p_a(\cdot)$ e sono inoltre indipendenti dalle variabili casuali $(Y_j)_{j=1}^{N^a(T)}$. Otteniamo quindi per il processo X_T^a la stessa legge del processo X^a in quanto viene costruito usando oggetti stocastici aventi le stesse leggi. Ossia abbiamo

$$\mathbb{P}(X_T^a(t) = b) = \mathbb{P}(X^a(t) = b) := p_t(a, b), \quad (1.3)$$

dove l'ultima uguaglianza e' la definizione del simbolo $p_t(a, b)$ che viene chiamata probabilit  di transizione da a a b nel tempo t . Si noti che evidentemente gode delle propriet : $p_t(a, b) \geq 0$ e

$$\sum_{b \in A} p_t(a, b) = 1, \quad \forall a \in A.$$

2 Propriet  di Markov e distribuzioni finito dimensionali

Cominciamo con la seguente definizione.

Definizione 2.1. Diremo che un processo stocastico X a valori in un insieme finito A ha la propriet  di Markov se succede che comunque scelti n tempi $t_1 < t_2 < \dots < t_n$ ed n stati a_1, \dots, a_n abbiamo che

$$\mathbb{P}\left(X(t_n) = a_n \mid X(t_{n-1}) = a_{n-1}, \dots, X(t_1) = a_1\right) = \mathbb{P}\left(X(t_n) = a_n \mid X(t_{n-1}) = a_{n-1}\right). \quad (2.1)$$

Facciamo ora vedere che una catena di Markov a tempo continuo soddisfa la propriet  di Markov.

Studiamo dapprima la seguente probabilit :

$$\mathbb{P}(X^a(t_1) = a_1, X^a(t_2) = a_2), \quad t_1 < t_2. \quad (2.2)$$

Possiamo disegnare nello schema della costruzione grafica due rette orizzontali in corrispondenza dei tempi t_1 e t_2 . In base alle regole grafiche per avere che $X^a(t_1) = a_1$ si devono verificare all'interno della finestra temporale $(0, t_1]$ delle combinazioni di pallini \bullet e frecce che portino un pallino \circ posto in a all'istante 0 fino allo stato a_1 al tempo t_1 . Questo corrisponde all'evento $\{X^a(t_1) = a_1\}$. Per avere inoltre anche $X(t_2) = a_2$ si dovr  avere una combinazione di pallini e frecce nella finestra temporale $(t_1, t_2]$ che porti una pallina \circ posta in a_1 all'istante t_1 fino ad a_2 all'istante t_2 . Questo corrisponde all'evento $\{X_{t_1}^{a_1}(t_2 - t_1) = a_2\}$.

Come osservato nella sezione precedente le costruzioni grafiche nelle finestre temporali $(0, t_1]$ e $(t_1, t_2]$ sono tra di loro indipendenti ed utilizzando (1.3) otteniamo che (2.2) e' uguale a

$$\begin{aligned} \mathbb{P}(X^a(t_1) = a_1, X^a(t_2) = a_2) &= \mathbb{P}(X^a(t_1) = a_1, X_{t_1}^{a_1}(t_2 - t_1) = a_2) \\ &= \mathbb{P}(X^a(t_1) = a_1) \mathbb{P}(X_{t_1}^{a_1}(t_2 - t_1) = a_2) = p_{t_1}(a, a_1) p_{t_2 - t_1}(a_1, a_2). \end{aligned} \quad (2.3)$$

Il ragionamento precedente puo' essere facilmente generalizzato ad n tempi ottenendo

$$\mathbb{P}(X^a(t_1) = a_1, \dots, X^a(t_n) = a_n) = p_{t_1}(a, a_1)p_{t_2-t_1}(a_1, a_2) \dots p_{t_n-t_{n-1}}(a_{n-1}, a_n). \quad (2.4)$$

Dalla formula (2.4) applicando la definizione di probabilità condizionata otteniamo

$$\begin{aligned} & \mathbb{P}\left(X^a(t_n) = a_n \mid X^a(t_{n-1}) = a_{n-1}, \dots, X^a(t_1) = a_1\right) \\ &= \frac{\mathbb{P}(X^a(t_n) = a_n, X^a(t_{n-1}) = a_{n-1}, \dots, X^a(t_1) = a_1)}{\mathbb{P}(X^a(t_{n-1}) = a_{n-1}, \dots, X^a(t_1) = a_1)} \\ &= \frac{p_{t_1}(a, a_1)p_{t_2-t_1}(a_1, a_2) \dots p_{t_n-t_{n-1}}(a_{n-1}, a_n)}{p_{t_1}(a, a_1)p_{t_2-t_1}(a_1, a_2) \dots p_{t_{n-1}-t_{n-2}}(a_{n-2}, a_{n-1})} \\ &= p_{t_n-t_{n-1}}(a_{n-1}, a_n) \\ &= \frac{p_{t_{n-1}}(a, a_{n-1})p_{t_n-t_{n-1}}(a_{n-1}, a_n)}{p_{t_{n-1}}(a, a_{n-1})} \\ &= \frac{\mathbb{P}(X^a(t_n) = a_n, X^a(t_{n-1}) = a_{n-1})}{\mathbb{P}(X^a(t_{n-1}) = a_{n-1})} \\ &= \mathbb{P}\left(X^a(t_n) = a_n \mid X^a(t_{n-1}) = a_{n-1}\right). \end{aligned}$$

ossia la validità della proprietà di Markov (2.1) per il processo X^a con inoltre l'identificazione

$$p_{t_n-t_{n-1}}(a_{n-1}, a_n) = P\left(X^a(t_n) = a_n \mid X^a(t_{n-1}) = a_{n-1}\right).$$

Si lascia come esercizio per il lettore la verifica della validità della proprietà di Markov per il processo X^μ con associata le medesime probabilità di transizione.

Osserviamo che (2.4) fornisce le distribuzioni finito dimensionali del processo X^a ; poichè deve valere la condizione di compatibilità otteniamo immediatamente le cosiddette equazioni di Chapman-Kolmogorov che affermano la seguente proprietà delle probabilità di transizione

$$p_{t+s}(a, c) = \sum_{b \in A} p_t(a, b)p_s(b, c), \quad t, s \geq 0. \quad (2.5)$$

3 Semigruppì e generatori

Consideriamo una funzione $f : A \rightarrow \mathbb{R}$ e definiamo per ogni $t \geq 0$ un'operatore lineare $S(t)$ che trasforma la funzione f nella funzione $S(t)f$ definita da

$$(S(t)f)(a) := \mathbb{E}(f(X^a(t))). \quad (3.1)$$

La linearità dell'operatore segue dalla linearità del valore di aspettazione. Notiamo che l'operatore $S(0)$ coincide con l'operatore identità \mathbb{I} . La definizione (3.1) può essere resa più esplicita osservando che vale

$$\mathbb{E}(f(X^a(t))) = \sum_{b \in A} p_t(a, b)f(b). \quad (3.2)$$

Facciamo ora vedere che l'equazione di Chapman-Kolmogorov (2.5) implica la validità della proprietà di semigrupp

$$S(t+s) = S(t)S(s), \quad s, t \geq 0. \quad (3.3)$$

Si parla di semigrupp e non di gruppo in quanto (3.3) e' valida solamente per t ed s positivi. Dimostriamo ora (3.3) applicando gli operatori ad una generica funzione f , otteniamo

$$\begin{aligned} (S(t+s)f)(a) &= \sum_{c \in A} p_{t+s}(a, c) f(c) \\ &= \sum_{c \in A} \sum_{b \in A} p_t(a, b) p_s(b, c) f(c) \\ &= \sum_{b \in A} p_t(a, b) \sum_{c \in A} p_s(b, c) f(c) \\ &= \sum_{b \in A} p_t(a, b) (S(s)f)(b) \\ &= (S(t)(S(s)f))(a). \end{aligned}$$

Osserviamo che una funzione a valori reali definita sull'insieme finito A può essere rappresentata in modo naturale come un vettore $f \in \mathbb{R}^{|A|}$: la componente del vettore f in corrispondenza con l'elemento a vale $f(a)$. E' quindi naturale rappresentare gli operatori lineari $S(t)$ come delle matrici $|A| \times |A|$ i cui elementi indicheremo con $S(t)_{a,b}$. Avremo quindi

$$(S(t)f)(a) = \sum_{b \in A} S(t)_{a,b} f(b).$$

Dalla formula (3.2) otteniamo inoltre

$$S(t)_{a,b} = p_t(a, b). \quad (3.4)$$

Naturalmente anche una misura di probabilità su A può essere rappresentata in modo naturale con un vettore e conseguentemente un operatore lineare agente sulle misure con una matrice.

L'evoluzione sulle funzioni reali definite su A che abbiamo appena definito determina in modo canonico una evoluzione sull'insieme delle misure di probabilità su A . Per definirla ricordiamo la dualità tra misure e funzioni che associa ad una coppia μ misura di probabilità e f funzione il numero reale

$$\mathbb{E}_\mu(f) = \sum_{a \in A} \mu(a) f(a),$$

dove con il simbolo \mathbb{E}_μ abbiamo indicato il valore di aspettazione rispetto alla misura μ . Possiamo quindi definire una famiglia di operatori lineari dipendenti dal parametro t che chiameremo $S^T(t)$ (chiariremo subito il motivo della scelta simbolica: il simbolo T starà ad indicare un'operazione di trasposizione) che agisce sulle misure di probabilità su A . La misura $S^T(t)\mu$ sarà definita dalla seguente relazione

$$\mathbb{E}_{S^T(t)\mu}(f) := \mathbb{E}_\mu(S(t)f), \quad (3.5)$$

dove f e' una generica funzione. Ricordiamo che la conoscenza dei valori di aspettazione rispetto ad una misura di una qualunque funzione determina completamente la misura stessa: ad esempio considerando $f(a) = \delta_{a,a^*}$ otteniamo

$$\mathbb{E}_\mu(f) = \mu(a^*).$$

Sviluppando (3.5) rappresentando gli operatori $S(t)$ ed $S^T(t)$ tramite le corrispondenti matrici otteniamo

$$\sum_{a \in A} \sum_{b \in A} S^T(t)_{a,b} \mu(b) f(a) = \sum_{a \in A} \sum_{b \in A} \mu(a) S(t)_{a,b} f(b).$$

Poichè la relazione precedente deve essere vera per ogni funzione f e per ogni misura di probabilità μ otteniamo

$$S^T(t)_{a,b} = S(t)_{b,a} = p_t(b, a),$$

ossia la matrice associata all'operatore $S^T(t)$ e' la trasposta della matrice associata all'operatore $S(t)$ e questo giustifica la notazione utilizzata.

Alcuni testi preferiscono rappresentare una funzione f come un vettore colonna ed una misura di probabilità come vettore riga e non introdurre l'operatore $S^T(t)$ bensì utilizzare le notazioni $S(t)f$ dove l'operatore $S(t)$ agisce a destra ed $\mu S(t)$ dove l'operatore $S(t)$ agisce a sinistra.

In definitiva comunque otteniamo

$$(S^T(t)\mu)(a) = \sum_{b \in A} S(t)_{b,a} \mu(b).$$

La misura $S^T(t)\mu$ e' ancora una misura di probabilità infatti

$$\sum_{a \in A} (S^T(t)\mu)(a) = \sum_{a \in A} \sum_{b \in A} S^T(t)_{a,b} \mu(b) = \sum_{b \in A} \mu(b) \sum_{a \in A} p_t(b, a) = 1$$

Inoltre abbiamo anche

$$(S^T(t)\mu)(a) = \sum_{b \in A} \mu(b) p_t(b, a) = \mathbb{P}(X^\mu(t) = a).$$

Ossia se la catena di Markov X ha distribuzione μ al tempo 0 allora avrà distribuzione $S^T(t)\mu$ al tempo t .

Vogliamo ora calcolare

$$\lim_{t \rightarrow 0} \frac{p_t(a, b)}{t} = \lim_{t \rightarrow 0} \frac{\mathbb{P}(X^a(t) = b)}{t}, \quad b \neq a. \quad (3.6)$$

Per le proprietà del processo di Poisson abbiamo che $\mathbb{P}(N(t) > 1) = o(t)$ quindi il limite (3.6) e' uguale a

$$\begin{aligned} & \lim_{t \rightarrow 0} \frac{\mathbb{P}(X^a(t) = b, N^a(t) = 1, N(t) = 1)}{t} \\ &= \lim_{t \rightarrow 0} \frac{p_a(b) e^{-t \sum_{c \in A} \lambda_c} \lambda_a t}{t} = \lambda_a p_a(b). \end{aligned} \quad (3.7)$$

Il risultato $\lambda_a p_a(b)$ del precedente calcolo viene spesso chiamato tasso (o rate in inglese) di salto dallo stato a allo stato b . Utilizzando (3.7) possiamo calcolare

$$\begin{aligned} & \lim_{t \rightarrow 0} \frac{\mathbb{E}(f(X^a(t))) - f(a)}{t} \\ & \lim_{t \rightarrow 0} \frac{\sum_{b \in A} p_t(a, b)(f(b) - f(a))}{t} \\ & = \sum_{b \in A} \lambda_a p_a(b)(f(b) - f(a)) := (Lf)(a). \end{aligned} \quad (3.8)$$

La formula (3.8) definisce un operatore lineare L sullo spazio delle funzioni su A detto *Generatore*. Rappresentandolo con una matrice da (3.8) otteniamo

$$\sum_{b \in A} L_{a,b} f(b) = \sum_{b \neq a} \lambda_a p_a(b) f(b) - \lambda_a (1 - p_a(a)) f(a),$$

per cui gli elementi di matrice sono

$$L_{a,b} = \begin{cases} \lambda_a p_a(b) & \text{if } a \neq b, \\ -\lambda_a (1 - p_a(a)) & \text{if } a = b. \end{cases} \quad (3.9)$$

Si noti che la somma degli elementi di ogni riga della matrice L da zero: $\sum_{b \in A} L_{a,b} = 0$.

In termini matriciali possiamo scrivere (3.8) come

$$\lim_{t \rightarrow 0} \frac{S(t) - \mathbb{I}}{t} = L.$$

Si noti che dalla proprietà di semigrupp (3.3) possiamo ottenere

$$\frac{d}{dt} S(t) = \lim_{\Delta \rightarrow 0} \frac{S(t + \Delta) - S(t)}{\Delta} = \lim_{\Delta \rightarrow 0} \frac{S(t)(S(\Delta) - \mathbb{I})}{\Delta} = S(t)L$$

ed analogamente

$$\frac{d}{dt} S(t) = \lim_{\Delta \rightarrow 0} \frac{S(t + \Delta) - S(t)}{\Delta} = \lim_{\Delta \rightarrow 0} \frac{(S(\Delta) - \mathbb{I})S(t)}{\Delta} = LS(t).$$

In particolare abbiamo ottenuto dall'esistenza del limite (3.6) la differenziabilità del semigrupp S . Data una funzione f (che ricordiamo può essere identificata con un vettore in $\mathbb{R}^{|A|}$) abbiamo quindi che il vettore $f_t := S(t)f$ risolve il sistema di equazioni differenziali

$$\frac{d}{dt} f_t = L f_t, \quad (3.10)$$

con in aggiunta la condizione iniziale $f_0 = f$. Abbiamo quindi un problema di Cauchy la cui unica soluzione è espressa dalla formula

$$f_t = e^{tL} f, \quad (3.11)$$

dove e^{tL} è l'esponenziale della matrice tL .

In modo analogo si può ricavare che la misura $\mu_t := S^T(t)\mu$ soddisfa l'equazione differenziale

$$\frac{d}{dt}\mu_t = L^T \mu_t, \quad (3.12)$$

con associata la condizione iniziale $\mu_0 = \mu$. Abbiamo quindi

$$\mu_t = e^{tL^T} \mu. \quad (3.13)$$

Una catena di Markov a tempo continuo viene spesso individuata assegnandone il generatore. Una volta assegnato il generatore risultano determinate le probabilità di transizione e quindi le famiglie finite dimensionali. Si noti invece che la conoscenza degli elementi di matrice del generatore non permette di ricostruire univocamente i valori dei parametri λ_a e $p_a(b)$; questo è possibile farlo una volta che si decida di fissare $p_a(a) = 0$ per ogni $a \in A$. Questa particolare scelta corrisponde alla cosiddetta rappresentazione canonica della catena di Markov.

4 Accoppiamento

Descriviamo ora una tecnica molto importante della teoria della probabilità che va sotto il nome di *Accoppiamento o Coupling*. La descriviamo nel contesto delle catene di Markov a tempo continuo assumendo per semplicità che $L_{a,b} > 0$ per ogni $a \neq b$.

In modo sintetico, la tecnica dell'accoppiamento consiste nella costruzione di due oggetti aleatori (nel nostro caso due catene di Markov) sullo stesso spazio di probabilità (ovviamente la costruzione dovrà essere opportunamente generata per poterne dedurre informazioni importanti).

Vogliamo costruire due catene di Markov aventi gli stessi tassi di salto (ossia lo stesso generatore) ma stati iniziali diversi. Ossia vogliamo costruire nello stesso spazio di probabilità i processi X^a ed X^b dove a e b sono due elementi diversi di A . Notiamo subito che la costruzione grafica descritta in sezione 1 è perfetta allo scopo. Basta infatti porre al tempo zero due pallini (distinguibili!) di tipo \circ uno in corrispondenza dello stato a ed uno in corrispondenza dello stato b e farli evolvere utilizzando la stessa struttura grafica. In questo caso lo spazio di probabilità comune nel quale definiamo i due processi è quello nel quale sono definiti i processi di Poisson indipendenti N e le variabili casuali Y .

La caratteristica fondamentale di questo particolare accoppiamento è che se indichiamo con

$$\tau^{a,b} := \inf \{t : X^a(t) = X^b(t)\}, \quad (4.1)$$

il primo tempo di incontro delle due catene, allora avremo anche $X^a(t) = X^b(t)$ per ogni $t \geq \tau^{a,b}$. Ossia una volta che i due processi si incontrano poi procedono di pari passo.

Una seconda caratteristica di questo accoppiamento è che è un accoppiamento Markoviano. Con questo intendiamo che se consideriamo il processo stocastico $Z^{a,b}(t) := (X^a(t), X^b(t))$ avente spazio degli stati $A \times A$, questo processo è Markoviano. Questa proprietà può essere dedotta dalla regola grafica di evoluzione esattamente come fatto per il singolo processo X . Procedendo sempre come per il singolo processo X è possibile anche calcolare il generatore associato al processo Markoviano Z .

Sia $F : A \times A \rightarrow \mathbb{R}$ una funzione definita sullo spazio degli stati di Z . Definiamo il semigruppoo $\mathcal{S}(t)$ che agisce su queste funzioni come

$$(\mathcal{S}(t)F)(a, b) := \mathbb{E}(F(Z^{a,b}(t))).$$

Otteniamo che il corrispondente generatore \mathcal{L} definito da

$$(\mathcal{L}F)(a, b) := \lim_{t \rightarrow 0} \frac{(\mathcal{S}(t)F)(a, b) - F(a, b)}{t},$$

ha la seguente espressione. Abbiamo che $(\mathcal{L}F)(a, b)$ e' uguale ad

$$\sum_{a' \in A} \lambda_a p_a(a') [F(a', b) - F(a, b)] + \sum_{b' \in A} \lambda_b p_b(b') [F(a, b') - F(a, b)]$$

quando $a \neq b$. Abbiamo invece che $(\mathcal{L}F)(a, a)$ e' uguale a

$$\sum_{a' \in A} \lambda_a p_a(a') [F(a', a') - F(a, a)]$$

quando $a = b$. Consideriamo la funzione $G : A \times A \rightarrow \mathbb{R}$ definita come

$$G(a, b) := \begin{cases} 1 & \text{se } a \neq b, \\ 0 & \text{se } a = b. \end{cases}$$

Chiaramente abbiamo

$$(\mathcal{S}(t)G)(a, b) = \mathbb{P}(X^a(t) \neq X^b(t)) = \mathbb{P}(\tau^{a,b} > t).$$

Inoltre abbiamo anche che

$$(\mathcal{L}G)(a, b) = \begin{cases} -\lambda_a p_a(b) - \lambda_b p_b(a) & \text{se } a \neq b, \\ 0 & \text{se } a = b. \end{cases}$$

Introduciamo ora

$$\lambda^* := \min_{a \neq b} \{\lambda_a p_a(b) + \lambda_b p_b(a)\} > 0, \quad (4.2)$$

il fatto che λ^* e' strettamente positivo segue dal fatto che e' ottenuto come minimo di un numero finito di numeri strettamente positivi. Se definiamo

$$H(a, b) := \begin{cases} -\lambda^* & \text{se } a \neq b, \\ 0 & \text{se } a = b, \end{cases}$$

abbiamo che $\mathcal{L}G \leq H$. Dalla definizione di generatore deriviamo che, per ogni $a \neq b$,

$$\begin{aligned} \frac{d}{dt} (\mathbb{P}(\tau^{a,b} > t)) &= \frac{d}{dt} (\mathcal{S}(t)G)(a, b) = \mathcal{S}(t)\mathcal{L}G(a, b) \\ &\leq \mathcal{S}(t)H(a, b) = -\lambda^* \mathbb{P}(\tau^{a,b} > t). \end{aligned} \quad (4.3)$$

La disuguaglianza e' stata ottenuta osservando che $\mathcal{S}(t)F_1 \leq \mathcal{S}(t)F_2$ per ogni t quando $F_1 \leq F_2$. Dal lemma di Gronwall e dal fatto che $\mathbb{P}(\tau^{a,b} > 0) = 1$ otteniamo

$$\mathbb{P}(\tau^{a,b} > t) \leq e^{-\lambda^* t}. \quad (4.4)$$

Osserviamo che allo stesso modo si può costruire un accoppiamento tra due catene X^μ ed X^ν che invece di partire da due stati fissati partono distribuite secondo due misure μ e ν . Per costruire l'accoppiamento basta distribuire al tempo zero i due pallini di tipo \circ secondo le leggi μ e ν indipendentemente (ma in linea di principio si potrebbero accoppiare anche le condizioni iniziali) e poi farli evolvere utilizzando la stessa struttura grafica. Anche in questo caso i due processi proseguiranno insieme una volta incontratisi. Chiamando $\tau^{\mu,\nu}$ il tempo di primo incontro dei processi accoppiati avremo

$$\mathbb{P}(\tau^{\mu,\nu} > t) = \sum_{a,b} \mu(a)\nu(a)\mathbb{P}(\tau^{a,b} > t) \leq e^{-\lambda^*t} \sum_{a,b} \mu(a)\nu(a) = e^{-\lambda^*t}.$$

5 Il problema delle misure invarianti

Data una catena di Markov a tempo continuo (ad esempio assegnato un generatore), un problema di interesse anche in vari contesti applicativi e' quello di trovarne le cosiddette *misure invarianti*. Una misura invariante e' una misura di probabilità π su A tale che se distribuiamo al tempo zero la catena di Markov secondo la legge π allora ad ogni tempo t avremo

$$\mathbb{P}(X^\pi(t) = a) = \pi(a).$$

In base ai risultati del paragrafo precedente abbiamo che π deve essere una soluzione costante del sistema di equazioni differenziali (3.12), ossia si deve avere

$$0 = L^T \pi, \tag{5.1}$$

che scritta in forma non compatta diviene

$$\sum_{b \neq a} \pi(b)L_{b,a} = \pi(a) \left(\sum_{b \neq a} L_{a,b} \right), \quad \forall a \in A. \tag{5.2}$$

Una naturale interpretazione dell'equazione (5.2) e' la seguente. Per ogni coppia $a \neq b \in A$ consideriamo il flusso $F_{a,b} := \pi(a)L_{a,b}$ uscente dallo stato a ed entrante nello stato b . Le equazioni (5.2) stabiliscono che per ogni elemento $a \in A$ si deve avere un perfetto bilancio tra flusso entrante e flusso uscente.

Poichè le equazioni (5.2) sono lineari, date $\pi^{(1)}$ e $\pi^{(2)}$ misure invarianti allora anche

$$c\pi^{(1)} + (1-c)\pi^{(2)}, \quad c \in [0, 1], \tag{5.3}$$

e' una misura invariante. Chiamiamo

$$\mathcal{M}_A := \left\{ \mu := (\mu(1), \dots, \mu(A)) : \mu(a) \geq 0, \sum_a \mu(a) = 1 \right\}$$

l'insieme di tutte le misure di probabilità su A . Da (5.3) deduciamo che l'insieme $\mathcal{I} \subseteq \mathcal{M}_A$ di tutte le misure invarianti e' un sottoinsieme convesso di \mathcal{M}_A .

Poichè A e' un insieme finito possiamo dimostrare che \mathcal{I} e' un insieme non vuoto, ma questo non e' vero in generale ed esistono catene di Markov con spazi degli stati infiniti che

non ammettono misure invarianti. Data una misura iniziale μ ricordiamo che indichiamo con $\mu_t = S^T(t)\mu$ la distribuzione del processo al tempo t . Definiamo

$$\nu_t := \frac{1}{t} \int_0^t \mu_s ds. \quad (5.4)$$

Poichè per ogni t abbiamo $\nu_t \in \mathcal{M}_A$ e poichè \mathcal{M}_A è compatto deduciamo che possiamo estrarre una sottosuccessione $\nu_n := \nu_{t_n}$ (t_n è una successione crescente e divergente $t_n \rightarrow +\infty$ di tempi indicizzati dai numeri naturali) convergente ad un qualche elemento $\nu^* \in \mathcal{M}_A$. Sottolineiamo che la compattezza e la convergenza in questo contesto possono essere pensate nella usuale metrica Euclidea di \mathbb{R}^A . Facciamo ora vedere che necessariamente $\nu^* \in \mathcal{I}$ e quindi $\mathcal{I} \neq \emptyset$. Questo segue immediatamente dal fatto che per ogni $h > 0$ abbiamo

$$S^T(h)\nu^* = \nu^*. \quad (5.5)$$

Infatti abbiamo

$$\begin{aligned} S^T(h)\nu^* &= S^T(h) \lim_{n \rightarrow +\infty} \nu_n \\ &= S^T(h) \lim_{n \rightarrow +\infty} \frac{1}{t_n} \int_0^{t_n} \mu_s ds \\ &= \lim_{n \rightarrow +\infty} \frac{1}{t_n} \int_0^{t_n} \mu_{s+h} ds \\ &= \lim_{n \rightarrow +\infty} \frac{1}{t_n} \int_h^{t_n+h} \mu_s ds \\ &= \lim_{n \rightarrow +\infty} \frac{1}{t_n} \left(\int_0^{t_n} \mu_s ds - \int_0^h \mu_s ds + \int_{t_n}^{t_n+h} \mu_s ds \right) \\ &= \lim_{n \rightarrow +\infty} \frac{1}{t_n} \int_0^{t_n} \mu_s ds = \nu^*. \end{aligned}$$

La penultima uguaglianza segue dal fatto che t_n è una sequenza divergente e quindi i due ultimi termini convergono a zero nel limite.

Sotto opportune condizioni è possibile dimostrare che la misura invariante è unica. Per semplicità ci limitiamo a dimostrare questo risultato di unicità nel caso in cui $L_{a,b} > 0$ per ogni $a \neq b$ ossia nel caso in cui si abbia un tasso di salto dal generico stato a al generico stato b strettamente positivo. Più in generale è possibile dimostrare l'unicità della misura invariante sotto l'ipotesi di irriducibilità della catena stessa, ossia sotto l'ipotesi che per ogni coppia di stati a, b esista un cammino di stati a_0, a_1, \dots, a_n tali che $a_0 = a$, ed $a_n = b$ ed inoltre $L_{a_i, a_{i+1}} > 0$ per ogni $i = 0, 1, \dots, n-1$.

Utilizziamo l'accoppiamento descritto nella sezione 4 e ipotizziamo per assurdo che esistano due diverse misure invarianti μ e ν . Costruiamo $Z^{\mu, \nu} = (X^\mu, X^\nu)$ il processo

accoppiato come nella sezione 4. Allora abbiamo

$$\begin{aligned}
\mu(a) &= \mathbb{P}(X^\mu(t) = a) \\
&= \mathbb{P}(X^\mu(t) = a | t \leq \tau^{\mu,\nu}) \mathbb{P}(t \leq \tau^{\mu,\nu}) + \mathbb{P}(X^\mu(t) = a | t > \tau^{\mu,\nu}) \mathbb{P}(t > \tau^{\mu,\nu}) \\
&= \mathbb{P}(X^\mu(t) = a | t \leq \tau^{\mu,\nu}) \mathbb{P}(t \leq \tau^{\mu,\nu}) + \mathbb{P}(X^\nu(t) = a | t > \tau^{\mu,\nu}) \mathbb{P}(t > \tau^{\mu,\nu}) \\
&= \mathbb{P}(X^\nu(t) = a) + \left(\mathbb{P}(X^\mu(t) = a | t < \tau^{\mu,\nu}) - \mathbb{P}(X^\nu(t) = a | t < \tau^{\mu,\nu}) \right) \mathbb{P}(t < \tau^{\mu,\nu}) \\
&= \nu(a) + \left(\mathbb{P}(X^\mu(t) = a | t < \tau^{\mu,\nu}) - \mathbb{P}(X^\nu(t) = a | t < \tau^{\mu,\nu}) \right) \mathbb{P}(t < \tau^{\mu,\nu}).
\end{aligned}$$

Poichè la catena di uguaglianze sopra illustrata e' valida per ogni t e poichè

$$\lim_{t \rightarrow +\infty} \mathbb{P}(t < \tau^{\mu,\nu}) = 0,$$

otteniamo che necessariamente $\mu = \nu$ e quindi la misura invariante e' unica.

6 Reversibilità

Consideriamo $\{X^\pi(t)\}_{t \in [-T, T]}$ una catena di Markov stazionaria osservata nell'intervallo temporale $[-T, T]$ (per poterla costruire ad esempio utilizzando la costruzione grafica si può procedere costruendo la catena nella finestra temporale $[0, 2T]$ e poi traslando il tutto di $-T$). Il termine stazionaria e' relativo al fatto che π e' una misura invariante della catena; una catena di Markov che ha come condizione iniziale una misura invariante viene detta catena stazionaria. Per una catena stazionaria si ha ad esempio che $\mathbb{P}(X^\pi(t) = a) = \pi(a)$ per ogni tempo t . Diremo che la catena di Markov X^π e' reversibile se il processo stocastico $\{Y(t)\}_{t \in [-T, T]}$ definito da

$$Y(t) := X^\pi(-t), \quad (6.1)$$

ha la stessa distribuzione di X^π . Notiamo che utilizzando questa definizione il processo Y ha traiettorie che non sono CADLAG. Esse infatti sono continue a sinistra e aventi limite da destra (CAGLAD). Possiamo immaginare di modificare il valore del processo Y in corrispondenza dei punti di discontinuità in modo da ottenere un processo con traiettorie CADLAG. E' facile vedere che tale operazione non cambia la distribuzione del processo. Da qui in seguito quindi penseremo il processo Y ottenuto tramite la trasformazione di inversione temporale (6.1) ed il cambiamento di valore nei punti di discontinuità che rende il processo un processo con traiettorie CADLAG.

Per avere quindi reversibilità, comunque fissati dei tempi $t_1, \dots, t_n \in [-T, T]$ e comunque fissati degli elementi $a_1, \dots, a_n \in A$, dobbiamo avere

$$\mathbb{P}(X^\pi(t_1) = a_1, \dots, X^\pi(t_n) = a_n) = \mathbb{P}(X^\pi(-t_1) = a_1, \dots, X^\pi(-t_n) = a_n). \quad (6.2)$$

Cominciamo con il far vedere che il processo Y e' anch'esso una catena di Markov a tempo continuo ed identifichiamone le probabilità di transizione ed il generatore. Le distribuzioni finito dimensionale sono

$$\begin{aligned}
\mathbb{P}(Y(t_1) = a_1, \dots, Y(t_n) = a_n) &= \mathbb{P}(X^\pi(-t_n) = a_n, \dots, Y(-t_1) = a_1) \\
&= \sum_a \pi(a) p_{T-t_n}(a, a_n) p_{t_n-t_{n-1}}(a_n, a_{n-1}) \dots p_{t_2-t_1}(a_2, a_1) \\
&= \pi(a_n) p_{t_n-t_{n-1}}(a_n, a_{n-1}) \dots p_{t_2-t_1}(a_2, a_1) \\
&= \pi(a_1) \tilde{p}_{t_2-t_1}(a_1, a_2) \dots \tilde{p}_{t_n-t_{n-1}}(a_{n-1}, a_n).
\end{aligned} \quad (6.3)$$

Nei passaggi precedenti abbiamo usato il fatto che π e' una misura invariante della catena, e quindi in particolare

$$\sum_a \pi(a) p_t(a, b) = \pi(b) \quad (6.4)$$

per ogni $t > 0$. Inoltre abbiamo chiamato

$$\tilde{p}_t(a, b) := p_t(b, a) \frac{\pi(b)}{\pi(a)}. \quad (6.5)$$

Si noti che \tilde{p}_t e' una probabilita' di transizione in quanto grazie a (6.4) vale la condizione di normalizzazione $\sum_b \tilde{p}_t(a, b) = 1$ per ogni a . Inoltre π e' una misura invariante per una catena di Markov con probabilita' di transizione \tilde{p} infatti

$$\sum_a \pi(a) \tilde{p}_t(a, b) = \sum_a \pi(a) p_t(b, a) \frac{\pi(b)}{\pi(a)} = \pi(b) \sum_a p_t(b, a) = \pi(b).$$

La formula (6.3) ci dice quindi che il processo Y e' una catena di Markov stazionaria avente probabilita' di transizione \tilde{p}_t . Il generatore del processo Y sarA quindi

$$\tilde{L}_{a,b} = \lim_{t \rightarrow 0} \frac{\tilde{p}_t(a, b)}{t} = \lim_{t \rightarrow 0} \frac{\pi(b) p_t(b, a)}{\pi(a) t} = \frac{\pi(b)}{\pi(a)} L_{b,a}, \quad a \neq b. \quad (6.6)$$

La condizione di reversibilita' (6.2) sarA verificata quando i generatori \tilde{L} e L delle catene Y ed X^π coincidono. Questo avviene quando

$$\pi(a) L_{a,b} = \pi(b) L_{b,a}, \quad a \neq b, \quad (6.7)$$

questa e' la condizione di reversibilita' per un generatore L e va sotto il nome di condizione di bilancio dettagliato. Dato un generatore L di una catena di Markov a tempo continuo per vedere se e' reversibile (pensiamo di essere nel caso di catene irriducibili aventi quindi un'unica misura invariante) bisogna cercare una misura di probabilita' π che soddisfi la condizione di bilancio dettagliato (6.7). Si noti che se una misura π soddisfa (6.7) allora essa e' necessariamente invariante per la catena con generatore L , infatti abbiamo

$$\sum_{b:b \neq a} \pi(b) L_{b,a} = \sum_{b:b \neq a} \pi(a) L_{a,b} = \pi(a) \sum_{b:b \neq a} L_{a,b},$$

che e' la condizione di stazionarieta' (5.2).

Terminiamo questa sezione dando una interpretazione astratta di quanto fatto. Indichiamo con $\langle \cdot, \cdot \rangle_\pi$ il prodotto scalare in $L^2(\pi)$, ossia date f e g due funzioni reali definite su A abbiamo

$$\langle f, g \rangle_\pi := \sum_a f(a) g(a) \pi(a).$$

Un generatore L puo' essere interpretato in modo naturale come un operatore in $L^2(\pi)$ che associa alla funzione f la funzione

$$Lf(a) := \sum_b L_{a,b} f(b) = \sum_{b:b \neq a} L_{a,b} (f(b) - f(a)).$$

Il generatore \tilde{L} del processo invertito nel tempo Y coincide con l'aggiunto in $L^2(\pi)$ di L , ossia $\tilde{L} = L^+$, verifichiamolo. L'operatore aggiunto L^+ è definito dalla condizione

$$\langle L^+ f, g \rangle_\pi = \langle f, Lg \rangle_\pi, \quad \forall f, g. \quad (6.8)$$

Il lato destro di (6.8) è infatti

$$\sum_a \pi(a) f(a) \sum_b L_{a,b} g(b),$$

cambiando nome agli indici muti di somma possiamo scriverlo come

$$\sum_a \pi(a) g(a) \sum_b \frac{\pi(b)}{\pi(a)} L_{b,a} f(b), \quad (6.9)$$

che deve essere uguale ad

$$\sum_a \pi(a) g(a) \sum_b L_{a,b}^+ f(b),$$

che è il lato sinistro di (6.8), per ogni scelta di f e g . Scegliendo $g(a) := \delta_{a,a^*}$ e $f(a) := \delta_{a,b^*}$ otteniamo

$$\pi(a^*) L_{a^*,b^*}^+ = \pi(b^*) L_{b^*,a^*},$$

che è la stessa condizione (6.6) ed identifica L^+ con \tilde{L} . Con questa interpretazione possiamo dire che un generatore è reversibile se e solo se è autoaggiunto in $L^2(\pi)$.

Sottolineiamo il fatto che la condizione di reversibilità dipende dalla particolare misura invariante π che si è scelta. Nel caso di unicità della misura invariante non ci sono ambiguità e si può parlare di reversibilità senza dover sottolineare rispetto a quale misura invariante si sta considerando il problema.

7 Approccio all'equilibrio

Anche in questa sezione ci mettiamo nell'ipotesi semplificativa $L_{a,b} > 0$ che come sappiamo garantisce l'unicità della misura invariante. Facciamo vedere che sotto queste ipotesi vale anche un risultato di ergodicità per la catena, ossia data una qualunque misura iniziale μ allora avremo che $\mu_t = S^T(t)\mu$ converge quando $t \rightarrow +\infty$ all'unica misura invariante π .

Essendo le misure di probabilità elementi di \mathbb{R}^A sarebbe naturale studiare la convergenza nel senso della metrica Euclidea, la studieremo invece secondo la metrica equivalente della distanza in variazione totale che viene spesso utilizzata in teoria della misura. Date μ e ν definiamo la loro distanza in variazione totale come

$$d_{TV}(\mu, \nu) := \frac{1}{2} \sum_{a \in A} |\mu(a) - \nu(a)|. \quad (7.1)$$

Lo scopo di questa sezione è quello di dimostrare che $d_{TV}(\mu_t, \pi)$ converge a zero quando $t \rightarrow +\infty$ e di dare anche una stima della velocità di convergenza. Anche per affrontare

questo problema utilizzeremo l'accoppiamento della sezione 4, in particolare consideriamo l'accoppiamento tra il processo X^μ ed il processo stazionario X^π . Abbiamo che

$$\begin{aligned}
 d_{TV}(\mu_t, \pi) &= \frac{1}{2} \sum_a |\mathbb{P}(X^\mu(t) = a) - \mathbb{P}(X^\pi(t) = a)| \\
 &= \frac{1}{2} \sum_a |\mathbb{P}(X^\mu(t) = a | t < \tau^{\mu, \pi}) - \mathbb{P}(X^\pi(t) = a | t < \tau^{\mu, \pi})| \mathbb{P}(t < \tau^{\mu, \pi}) \\
 &\leq e^{-\lambda^* t} \frac{1}{2} \sum_a (\mathbb{P}(X^\mu(t) = a | t < \tau^{\mu, \pi}) + \mathbb{P}(X^\pi(t) = a | t < \tau^{\mu, \pi})) \\
 &= e^{-\lambda^* t}.
 \end{aligned}$$

Il risultato qui sopra ottenuto viene sintetizzato nelle parole: la catena di Markov converge all'equilibrio con un tasso esponenziale.