21. Proprietà ergodica

Sia $(X_n)_n$ una catena di Markov omogenea con spazio degli stati finito o numerabile. Supponiamo che la sua matrice di transizione $\mathbf{P}$ sia irriducibile. Sia $\overline{T}_{jj}$ il tempo di ritorno atteso nello stato j e sia $V_j^{(n)}(x)$ il numero di visite nello stato j fatte dal cammino x nei primi n passi. In parole povere il *teorema ergodico* afferma che al crescere del numero di passi n, la frequenza delle visite nello stato j converge quasi certamente a $\frac{1}{\overline{T}_{jj}}$,

$$\frac{V_j^{(n)}(x)}{n} \longrightarrow \frac{1}{\overline{T}_{jj}} \qquad \mathbb{P}\text{-quasi certamente,}$$

vedi Paragrafo 21.b. Discutiamo quindi brevemente di una notevole e importante applicazione del teorema ergodico, nota come metodo catena di Markov Montecarlo, utile al calcolo approssimato iterato di integrali.

21.a Numero di passi tra due visite consecutive

Sia $(X_n)_n$ una catena di Markov omogenea sullo spazio di probabilità $(\Omega, \mathscr{E}, \mathbb{P})$ con spazio degli stati S finito o numerabile e sia j uno stato: $j \in S$. Siano $\Omega_i := \{x \in \Omega \,|\, X_0(x) = j\}$ e $\mathbb{P}_j(A) := \mathbb{P}(A \,|\, X_0 = j)$, $A \in \mathscr{E}$ la misura di probabilità condizionata da $X_0 = j$. Indichiamo con f_{ij} la probabilità di visitare j partendo da i e con $\overline{T}_{jj}$ il tempo medio di ritorno in j.

A partire dalla successione dei tempi di passaggio $\left(T_j^{(r)}\right)_r$ in j, vedi (20.8), possiamo definire le variabili aleatorie $\left(S_j^{(r)}(x)\right)_r$ $r = 1, 2, \dots$, dove $S_j^{(1)} = T_j^{(1)}$ e, per $r \geq 2$,

$$S_j^{(r)}(x) := \begin{cases} T_j^{(r)}(x) - T_j^{(r-1)}(x) & \text{se } T_j^{(r-1)}(x) < +\infty, \\ +\infty & \text{altrimenti.} \end{cases}$$

Ciascuna variabile aleatoria $S_j^{(r)}(x)$ conta il numero di passi tra la visita $(r-1)$-esima e la visita r-esima nello stato j.

21.1 Teorema. *Sia $j \in S$ uno stato ricorrente e sia $i \in S$ tale che $f_{ij} = 1$. Allora*

(i) $S_j^{(r)}(x)$ *è finito per quasi ogni* $x \in \Omega_i$ *e per ogni* $r \geq 1$.

(ii) $\left(S_j^{(r)}\right)_r$ *è una successione di variabili aleatorie indipendenti rispetto alla probabilità* $\mathbb{P}_i$.

(iii) *Le variabili aleatorie* $S_j^{(r)}$, $r \geq 2$, *sono identicamente distribuite rispetto alla probabilità* $\mathbb{P}_i$ *e, per ogni* $n \geq 1$ *e* $r \geq 2$,

$$\mathbb{P}_i(S_j^{(r)} = n) = f_{jj}^{(n)} \qquad \text{e} \qquad \mathbb{E}_i\left[S_j^{(r)}\right] = \overline{T}_{jj}.$$

Dimostrazione. (i) Poiché j è ricorrente, ogni cammino da i visita j infinite volte con probabilità f_{ij},

$$\mathbb{P}\left(T_j^{(r)} < \infty \,\middle|\, X_0 = i\right) = f_{ij} = 1 \qquad \forall r \geq 1,$$

vedi Proposizione 20.5. Dunque per ogni $r \geq 2$, $T_j^{(r)}(x) < +\infty$ per q.o. $x \in \Omega_i$. In particolare, $1 \leq S_j^{(r)}(x) < +\infty$ per quasi ogni $x \in \Omega_i$. Quindi

$$\mathbb{P}_i(S_j^{(r)} = +\infty) = 0.$$

(ii) Siano $r \geq 2$, $h_1, \dots, h_{r-1} \geq 1$, $h = (h_1, \dots, h_{r-1}) \in \mathbb{N}^{r-1}$ e $n \geq 1$. Poniamo $k := h_1 + \cdots + h_{r-1}$ e siano $T(x) := T_j^{(r-1)}(x)$ e

$$H_h := \left\{x \in \Omega \,\middle|\, S_j^{(r-1)}(x) = h_{r-1},\ S_j^{(r-2)}(x) = h_{r-2}, \dots,\ S_j^{(1)}(x) = h_1\right\}.$$

L'evento H_h è rilevato dalle variabili aleatorie $X_0, \dots, X_k$ e $H_h = \{x \in \Omega \,|\, X_k = j\} \cap F_h$ dove F_h è rilevato da $X_0, \dots, X_{k-1}$. Si ha

$$\mathbb{P}_i\left(S_j^{(r)} = n \,\middle|\, H_h\right) = \frac{\mathbb{P}\left(\{S_j^{(r)} = n\} \cap H_h \,\middle|\, X_0 = i\right)}{\mathbb{P}\left(H_h \,\middle|\, X_0 = i\right)}$$

e poiché

$$\begin{aligned}&\left\{x \in \Omega \,\middle|\, S_j^{(r)} = n\right\} \cap H_h \\ &\qquad = \left\{x \in \Omega \,\middle|\, X_{n+k} = j, X_{n+k-1} \neq j, \dots, X_{k-1} \neq j, X_k = j\right\} \cap H_h\end{aligned}$$

si ha, ricordando che $H_h = \{x \in \Omega \,|\, X_k = j\} \cap F_h$ e utilizzando la proprietà di Markov e l'omogeneità della catena

$$
\begin{aligned}
\mathbb{P}\left(\{S_j^{(r)} = n\} \cap H_h \,\middle|\, X_0 = i\right)
&= \mathbb{P}\left(X_{n+k} = j,\, X_{n+k-1} \neq j, \dots,\, X_{k-1} \neq j \,\middle|\, \{X_k = j\} \cap F_h\right) \\
&\qquad \cdot \mathbb{P}\left(H_h \,\middle|\, X_0 = i\right) \\
&= \mathbb{P}\left(X_n = j,\, X_{n-1} \neq j, \dots,\, X_1 \neq j \,\middle|\, \{X_0 = j\} \cap F_h\right) \\
&\qquad \cdot \mathbb{P}\left(H_h \,\middle|\, X_0 = i\right) \\
&= f_{jj}^{(n)}\, \mathbb{P}\left(H_h \,\middle|\, X_0 = i\right).
\end{aligned}
$$

Ne segue che per ogni $r \geq 2$

$$\mathbb{P}_i\left(S_j^{(r)} = n \,\middle|\, H_h\right) = f_{jj}^{(n)}.$$

In particolare gli eventi individuati da $S_j^{(r)}$ sono indipendenti dagli eventi H_h, $h \in \mathbb{N}^{r-1}$, cioè indipendenti dagli eventi individuati da $S_j^{(r-1)}, \dots, S_j^{(1)}$. Dunque $S_j^{(r)}$ è indipendente da $S_j^{(r-1)}, \dots, S_j^{(1)}$ e, poiché r è arbitrario, $\left(S_j^{(r)}\right)_r$ è una successione di variabili aleatorie indipendenti rispetto alla misura $\mathbb{P}_i$.
(iii) Poiché gli H_h sono a due a due disgiunti e

$$\bigcup_{h \in \mathbb{N}^{r-1}} H_h = \left\{x \in \Omega \,\middle|\, S_j^{(r)} < +\infty\right\},$$

$\sum_{h \in \mathbb{N}^{r-1}} \mathbb{P}_i(H_h) = 1$. Pertanto da (ii) segue per ogni $r \geq 2$ e ogni $n \geq 1$

$$\mathbb{P}_i(S_j^{(r)} = n) = \sum_h \mathbb{P}_i\left(S_j^{(r)} = n \,\middle|\, H_h\right) \mathbb{P}_i(H_h) = f_{jj}^{(n)} \sum_{h \in \mathbb{N}^{r-1}} \mathbb{P}_i(H_h) = f_{jj}^{(n)} :$$

le variabili $\left(S_j^{(r)}\right)_{r \geq 2}$ sono identicamente distribuite rispetto a $\mathbb{P}_i$. Infine,

$$\mathbb{E}_i\left[S_j^{(r)}\right] = \sum_{k=0}^{\infty} k\, \mathbb{P}_i(S_j^{(r)} = k) = \sum_{k=1}^{\infty} k f_{jj}^{(k)} = \overline{T}_{jj}.$$

□

21.b Teorema ergodico

21.2 Proposizione. *Sia $(X_n)_n$ una catena di Markov omogenea con spazio degli stati S finito o numerabile. Per ogni $j \in S$ sia $\overline{T}_{jj}$ il tempo medio di ritorno in j e $w_j = 1/\overline{T}_{jj}$. Siano $i, j \in S$, $\Omega_i := \{x \in \Omega \,|\, X_0(x) = i\}$ e $\mathbb{P}_i(A) := \mathbb{P}(A \,|\, \Omega_i)$. Se*

j è transiente o se j è ricorrente e $f_{ij}=1$, allora la frequenza delle visite in j nei primi n passi, $\frac{V_j^{(n)}(x)}{n}$, converge a w_j per q.o. $x \in \Omega_i$, i.e.,

$$\frac{V_j^{(n)}(x)}{n} \to w_j \qquad \mathbb{P}_i\text{-quasi ovunque.} \tag{21.1}$$

Dimostrazione. Se lo stato j è transiente, per (20.16) $w_j = 0$ e i cammini da i quasi certamente visitano j solo un numero finito di volte, vedi Proposizione 17.7. Dunque

$$\frac{V_j^{(n)}(x)}{n} \le \frac{V_j(x)}{n} \to 0 = w_j \qquad \mathbb{P}\text{-quasi certamente.}$$

Se lo stato j è ricorrente e $f_{ij}=1$, si applica il Teorema 21.1. Con le notazioni del Teorema 21.1, le variabili aleatorie $S_j^{(r)}$, $r \ge 2$, sono indipendenti ed identicamente distribuite rispetto alla probabilità $\mathbb{P}_i$ con $\mathbb{E}_i\left[S_j^{(r)}\right] = \overline{T}_{jj} \in \overline{\mathbb{R}}_+$. Pertanto dalla legge dei grandi numeri, vedi Esercizio 14.23,

$$\frac{1}{n}\sum_{r=1}^{n} S_j^{(r)}(x) \ \to\ \overline{T}_{jj} \qquad \mathbb{P}_i\text{-quasi certamente.}$$

Poiché per $n \ge T_j^{(1)}(x) + S_j^{(2)}(x) = T_j^{(2)}(x)$

$$V_j^{(n)}(x) = \sup\left\{k \ge 2 \,\middle|\, \sum_{r=2}^{k} S_j^{(r)}(x) \le n\right\},$$

segue dalla Proposizione 15.8 che

$$\frac{V_j^{(n)}(x)}{n} \to \frac{1}{\overline{T}_{jj}} = w_j \qquad \mathbb{P}_i\text{-quasi certamente.}$$

$\square$

21.3 Teorema (Teorema ergodico). *Sia $(X_n)_n$ una catena di Markov omogenea con spazio degli stati S finito o numerabile e con matrice di transizione* **P** *irriducibile. Per ogni $j \in S$ siano $\overline{T}_{jj}$ il tempo medio di ritorno in j e $w_j = 1/\overline{T}_{jj}$. Allora*

(i) *Si ha*
$$\frac{V_j^{(n)}(x)}{n} \to w_j \qquad \mathbb{P}\text{-quasi certamente.} \tag{21.2}$$

(ii) *Sia $\phi : S \to \mathbb{R}$ una funzione tale che $\sum_{j\in S}|\phi(j)| < \infty$. Allora*
$$\frac{1}{n}\sum_{k=1}^{n} \phi(X_k) \ \to\ \sum_{j\in S}\phi(j)w_j \qquad \mathbb{P}\text{-quasi certamente.} \tag{21.3}$$

Dimostrazione. (i) Gli stati di S o sono tutti ricorrenti o sono tutti transienti e, se $j \in S$ è ricorrente, allora $f_{ij} = 1$, vedi Proposizione 20.5. Pertanto dalla Proposizione 21.2 per ogni $i \in S$, $\frac{V_j^{(n)}(x)}{n} \to w_j$ quasi certamente in $\Omega_i := \{x \in \Omega \,|\, X_0(x) = i\}$, quindi $\frac{V_j^{(n)}(x)}{n} \to w_j$ per $\mathbb{P}$-quasi ogni $x \in \Omega$.
(ii) Per $j \in S$ e $k \geq 1$, sia $E_{j,k} := \{x \in \Omega \,|\, X_k(x) = j\}$. Poiché

$$V_j^{(n)}(x) = \sum_{k=1}^{n} \mathbb{1}_{E_{k,j}}(x) \qquad \text{e} \qquad \phi(X_k(x)) = \sum_{j \in S} \phi(j)\, \mathbb{1}_{E_{j,k}}(x)$$

si ha

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n} \phi(X_k(x)) &= \frac{1}{n}\sum_{k=1}^{n}\Big(\sum_{j\in S}\phi(j)\,\mathbb{1}_{E_{j,k}}(x)\Big)\\ &= \sum_{j\in S}\phi(j)\frac{1}{n}\sum_{k=1}^{n}\mathbb{1}_{E_{j,k}}(x) = \sum_{j\in S}\phi(j)\frac{V_j^{(n)}(x)}{n}.\end{aligned}$$

Poiché $\sum_{j\in S}|\phi(j)| < \infty$, si può passare al limite con il teorema di convergenza dominata Esempio C.12, e ottenere

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\phi(X_k(x)) &= \lim_{n\to\infty}\sum_{j\in S}\phi(j)\frac{V_j^{(n)}(x)}{n} = \sum_{j\in S}\phi(j)\Big(\lim_{n\to\infty}\frac{V_j^{(n)}(x)}{n}\Big)\\ &= \sum_{j\in S}\phi(j) w_j \qquad \mathbb{P}\text{-quasi certamente.}\end{aligned}$$

□

21.4 Teorema. *Sia $(X_n)_n$ una catena di Markov omogenea con spazio degli stati S finito o numerabile e con matrice di transizione* $\mathbf{P}$. *Per ogni $j \in S$ sia $\overline{T}_{jj}$ il tempo medio di ritorno in j e $w_j = 1/\overline{T}_{jj}$. Per ogni $i, j \in S$ si ha*

$$\frac{1}{n}\sum_{k=1}^{n} p_{jj}^{(k)} \to w_j, \qquad \frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(k)} \to f_{ij} w_j. \tag{21.4}$$

Se $\mathbf{P}$ *è irriducibile, allora*

$$\frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(k)} \to w_j. \tag{21.5}$$

Dimostrazione. Dalla Proposizione 21.2 segue in particolare che $\frac{V_j^{(n)}(x)}{n} - w_j \to 0$ $\mathbb{P}_j$-quasi certamente. Inoltre $\frac{V_j^{(n)}(x)}{n} \leq 1$. Si applica il teorema di convergenza dominata, Teorema C.10, e si ottiene che

$$\int_{\Omega_j}\Big|\frac{V_j^{(n)}(x)}{n} - w_j\Big|\,\mathbb{P}(dx) \to 0,$$

in particolare

$$\mathbb{E}_j\left[V_j^{(n)}\right] = \frac{1}{\mathbb{P}(X_0 = j)} \int_{\{X_0=j\}} \frac{V_j^{(n)}(x)}{n}\, \mathbb{P}(dx) \to w_j$$

e quindi la prima delle (21.4), poiché

$$\frac{1}{n}\sum_{k=1}^{n} p_{jj}^{(k)} = \mathbb{E}_j\left[\frac{V_j^{(n)}}{n}\right].$$

La seconda delle (21.4) segue quindi dal Lemma 20.12.

Infine se $\mathbf{P}$ è irriducibile, allora la (21.5) segue dalla (21.4) poiché $w_j = 0$ se j è transiente e $f_{ij} = 1$ se j è ricorrente, vedi Proposizione 20.5. □

21.5 Corollario. *Sia* $(X_n)_n$ *una catena di Markov omogenea con spazio degli stati* $S = \{1,\dots,N\}$ *finito e con matrice di transizione* $\mathbf{P}$ *irriducibile. Per ogni* $j \in S$ *sia* $\overline{T}_{jj}$ *il tempo medio di ritorno in* j *e* $w_j = 1/\overline{T}_{jj}$*. Allora*

$$\frac{1}{n}\sum_{k=1}^{n} p_{ij}^{(k)} \to w_j,$$

$w := (w_1,\dots,w_N)$ *è l'unico vettore stocastico soluzione di* $w = w\mathbf{P}$ *e* $w_j > 0\ \forall j$*, i.e., ogni stato è positivamente ricorrente.*

Dimostrazione. Segue dal Teorema 21.4, dalla Proposizione 18.15 e dall'Esercizio 18.19. □

21.c Catena di Markov Monte Carlo

In molte applicazioni è necessario valutare la *somma pesata*, rispetto ad un peso $\pi = (\pi_j)$, di una certa quantità $f : S \to \mathbb{R}$ definita su un insieme finito S

$$E := \frac{1}{z}\sum_{j\in S} f(j)\pi_j, \qquad z := \sum_{j\in S}\pi_j. \tag{21.6}$$

Usando il linguaggio della probabilità, indichiamo con μ la probabilità su S con densità discreta $\frac{1}{z}(\pi_j)_{j\in S}$. La quantità da valutare dunque non è altro che

$$E := \int_S f(t)\mu(dt). \tag{21.7}$$

Se l'insieme S è piccolo, allora è possibile calcolare E direttamente. Se S è "grande", (per esempio $|E| = (50)!$), allora il calcolo di E o anche solo della massa totale z non può essere effettuato direttamente. È perciò opportuno avere almeno una strategia per calcolare, o quantomeno approssimare, il valore di E.

21.d Monte Carlo

Una di queste strategie è il cosiddetto metodo Monte Carlo, vedi Paragrafo 15.b. Si deve specificare uno spazio di probabilità $(\Omega, \mathscr{E}, \mathbb{P})$ e una variabile aleatoria $X : \Omega \to S$ con distribuzione $\mathbb{P}_X = \mu$; si considera poi lo schema di Bernoulli delle successioni a valori in Ω, $(\Omega^\infty, \mathscr{E}^\infty, \mathbb{P}^\infty)$, e le variabili aleatorie $X_n : \Omega^\infty \to \mathbb{R}$ definite, per ogni $n = 1, 2, \dots$, e $x = (x_n)_n \in \Omega^\infty$, come $X_n(x) := X(x_n)$. Le X_n costituiscono una successione di variabili aleatorie indipendenti ed identicamente distribuite come X, $\mathbb{P}_{X_n} = \mathbb{P}_X = \mu$, vedi Corollario 13.22. La legge forte dei grandi numeri afferma che, per quasi ogni successione di punti $x = (x_n)_n$, $x_n \in \Omega$,

$$\frac{1}{n}\sum_{j=1}^{n} f(X(x_j)) = \frac{1}{n}\sum_{j=1}^{n} f(X_j(x)) \;\to\; E \tag{21.8}$$

e vale la stima debole (14.3).

Per tradurre questo risultato in un algoritmo è necessario affrontare due questioni:

(i) Come scegliere (il più efficacemente possibile) lo spazio di probabilità $(\Omega, \mathscr{E}, \mathbb{P})$ e la variabile aleatoria $X : \Omega \to S$ avente distribuzione $\mathbb{P}_X = \mu$;
(ii) Poiché il limite (21.8) è vero solo per quasi ogni successione $(x_k)_k \in \Omega^\infty$ (e non per ogni successione), bisogna assicurarsi di selezionare una successione "buona". Il "rischio" di selezionare una "cattiva" successione è zero e tuttavia le "cattive" successioni esistono.
Se nella successione è presente una qualche forma di periodicità (o di regolarità), questa può influenzare il valore del limite o la sua stessa esistenza. Per esempio, supponiamo di lanciare una moneta equa infinite volte. Se si ottiene un successo (testa) ad ogni lancio o una volta ogni tre lanci, allora $\frac{1}{n}\sum_{k=1}^n X_k(x) = \frac{1}{x}\sum_{k=1}^n X(x_k)$ converge a 1 nel primo caso e a $1/3$ nel secondo e non al valore atteso $1/2$. Dovremmo metterci nelle condizioni di selezionare una successione il più possibile "irregolare" o "casuale".

Esistono molti modi per ottenere successioni o, meglio, vettori di "grande dimensione" di numeri casuali uniformemente distribuiti sull'intervallo $[0,1]$. È necessario essere cauti con la terminologia poiché questi numeri sono in realtà multipli interi di un numero fisso molto "piccolo". Tuttavia, se f non è una funzione fortemente oscillante e se l'approssimazione non deve essere estremamente accurata, questo fatto non crea problemi. Per esempio, *generatori fisici di numeri casuali* possono essere basati su fenomeni fisici atomici o subatomici la cui casualità è data dalle leggi della meccanica quantistica, quali il decadimento radioattivo, il rumore termico e altri. Con opportuni condizionamenti questi fenomeni sono sorgenti di campioni casuali uniformemente distribuiti sull'intervallo $[0,1]$.

Un'altra fonte utile sono i *generatori di numeri pseudo-casuali*. Si tratta di algoritmi deterministici che possono generare lunghe sequenze di numeri con "buone" proprietà stocastiche. Una casualità completa è fuori questione in quanto l'output di questi algoritmi è dato da sequenze periodiche, sebbene di periodo molto lungo. Inoltre le correlazioni tra output di indici molto diversi o tra gruppi di output vanno accuratamente controllate. Esistono comunque librerie di software che contengono generatori pseudo-casuali con proprietà statistiche certificate. Non ci occupiamo di questi aspetti che sono comunque rilevanti nel campo delle simulazioni, vedi per esempio [22], ma che non riguardano strettamente la probabilità. Per proseguire questa discussione, assumeremo che esistano generatori di "successioni casuali" uniformemente distribuite sull'intervallo $[0,1]$ e che sia possibile raggrupparli in N-uple in modo da ottenere una successione di N-uple uniformemente distribuite su $Q=[0,1]^N$.

Sulla base delle considerazioni precedenti, scelte tipiche per $(\Omega,\mathscr{E},\mathbb{P})$ e X per il calcolo di E nell'equazione (21.7) per $\mu=\mathscr{L}^N$ sono $(Q,\mathscr{B}(\mathbb{R}^N),\mathscr{L}^N)$ e $X(x):=x$.

Approssimare l'integrale (21.7) mediante la procedura (21.8) quando la misura μ non è uniforme, è cosa più complessa. Diamo un accenno alla procedura nel caso scalare. Supponiamo di voler calcolare

$$E:=\int_{-\infty}^{+\infty} f(t)\,d\mu(t)$$

con f funzione limitata e di Borel e $(\mathscr{B}(\mathbb{R}),\mu)$ misura su $\mathbb{R}$. Per prima cosa ci procuriamo una variabile aleatoria X con distribuzione μ. Per questo, si considera la legge $F(t):=\mu(]-\infty,t])$ di μ e si definisce la variabile aleatoria $X:[0,1]\to\mathbb{R}$ sullo spazio probabilizzato $([0,1],\mathscr{B}(\mathbb{R}),\mathscr{L}^1)$ come nella (9.13), cioè si pone

$$X(s):=\min\{t\,|\,s\le F(t)\}. \tag{21.9}$$

Poiché $F(t)$ è continua da destra, la $X(s)$ è continua da sinistra e dunque è di Borel. Inoltre $\{s\in[0,1]\,|\,X(s)\le t\}=\{s\in[0,1]\,|\,0\le s\le F(t)\}$ per ogni $t\in\mathbb{R}$ e dunque

$$\mathscr{L}^1(X\le t)=\mathscr{L}^1([0,F(t)])=F(t)\qquad\forall t\in\mathbb{R},$$

cioè la distribuzione di X è μ. Pertanto il metodo Monte Carlo afferma che per quasi ogni successione $(x_k)_k$ in $[0,1]$ "distribuita uniformemente" in $[0,1]$

$$\frac{1}{n}\sum_{k=1}^n f(X(x_k))\to\int_{-\infty}^{\infty} f(t)\,\mu(dt)\qquad \text{per } n\to\infty.$$

La procedura ora descritta non può essere implementata per valutare (21.6) quando l'insieme S è grande, poiché il calcolo della legge $F(t):=\mu(]-\infty,t])$ di μ o della variabile aleatoria X definita in (21.9) sono dello stesso ordine di complessità del calcolo diretto di E.

Quando l'insieme S può essere visto come un prodotto cartesiano, $S=S_1\times\dots\times S_N$ e la misura μ con densità di massa $\frac{1}{z}(\pi_j)$ è un prodotto,

$$\mu=\mu_1\times\mu_2\times\ldots\times\mu_N$$

dove ciascuna μ_i è una misura su S_i, è possibile fattorizzare il calcolo di X. Per esempio, se $N=2$, definiamo due variabile aleatorie $X:[0,1]\to S_1$, $Y:[0,1]\to S_2$ che seguono, rispettivamente, le distribuzioni μ_1 e μ_2. Allora

$$\int f(t,s)\,\mu_1\times\mu_2(dtds) = \int\Big(\int f(t,s)\mu_2(ds)\Big)\mu_1(dt)$$
$$= \int\Big(\int f(t,Y(y))\,\mathbb{P}(dy)\Big)\mu_1(dt) = \iint g(X(x),Y(y))\;\mathbb{P}\times\mathbb{P}(dxdy),$$

cosicché la legge forte dei grandi numeri dà

$$\frac{1}{n}\sum_{k=1}^{n} f(X(x_k),Y(y_k)) \to E$$

per quasi ogni successione $((x_k,y_k))_k$ di punti in $[0,1]^2$.

21.e Il metodo della catena di Markov Monte Carlo

Se μ non è una misura prodotto, è possibile sfruttare la proprietà ergodica delle catene di Markov.

Il paradigma che seguiamo per il calcolo di E nella (21.6) è il seguente: data la densità di probabilità $w = (w_j)_{j\in S}$, $w_j := \frac{1}{z}\pi_j$, definiamo una catena di Markov omogenea su uno spazio di probabilià $(\Omega,\mathscr{E},\mathbb{P})$ con spazio degli stati S e tale che w sia un punto fisso per la sua matrice di transizione $\mathbf{P}$, $w = w\mathbf{P}$, cioè tale che $\pi = \pi\mathbf{P}$.

Abbiamo già provato, vedi Teorema 19.9, che data una matrice stocastica $\mathbf{P}$ è sempre possibile definire una catena di Markov omogenea con matrice di transizione $\mathbf{P}$. Resta dunque da esibire una matrice stocastica $\mathbf{P}$ per la quale π sia un punto fisso: l'*algoritmo di Hastings–Metropolis* fornisce una matrice con questa proprietà.

Sia $\mathbf{Q} = (q_{ij})$ una matrice stocastica. Per $i,j\in S$, $i\neq j$, sia

$$p_{ij} := \frac{1}{\pi_i}\min\{\pi_i q_{ij},\pi_j q_{ji}\}, \qquad (21.10)$$

cosicché $0\le p_{ij}\le 1$ e $\sum_{j\neq i} p_{ij}\le 1$. Definiamo $\mathbf{P} = (p_{ij})$ come segue:

- ◦ se $i\neq j$, allora p_{ij} è definita dalla (21.10)
- ◦ se $i = j$, poniamo $p_{ii} = 1-\sum_{j\neq i} p_{ij}$,

di modo che $\mathbf{P}$ è una matrice stocastica. La matrice $\mathbf{R} = (r_{ij})$,

$$r_{ij} := \min\{\pi_i q_{ij},\pi_j q_{ji}\}$$

è simmetrica quindi

$$\pi_i p_{ij} = \pi_j p_{ji} \qquad \forall i,j\in S$$

e

$$\sum_{i=1}^{N}\pi_i p_{ij} = \sum_{i=1}^{N}\pi_j p_{ji} = \pi_j \qquad \forall j\in S$$

cioè $\pi = \pi\mathbf{P}$.

Nell'*algoritmo di Metropolis* si sceglie una matrice $\mathbf{Q} = (q_{ij})$ irriducibile e simmetrica, e per $i \neq j$ si pone

$$p_{ij} = \min\left\{1, \frac{\pi_j}{\pi_i}\right\} q_{ij},$$

e si definisce $\mathbf{P} = (p_{ij})$ come

$$p_{ij} = \begin{cases} q_{ij} & \text{se } \pi_j/\pi_i > 1, \\ \dfrac{\pi_j}{\pi_i} q_{ij} & \text{se } \pi_j < \pi_i, \\ 1 - \sum_{j \neq i} p_{ij} & \text{se } i = j. \end{cases} \tag{21.11}$$

Possibili scelte per la matrice $\mathbf{Q}$ sono

$$q_{ij} = \begin{cases} 1/(N-1) & \text{se } i \neq j, \\ 0 & \text{se } i = j, \end{cases}$$

dove $N = |S|$, o, per esempio si può scegliere come $\mathbf{Q}$ la matrice associata ad una passeggiata aleatoria simmetrica circolare. Osserviamo che in generale, per definire $\mathbf{P}$ non è necessario conoscere la massa totale z o la cardinalità N di S.

Sia $(X_n)_n$ una catena di Markov omogenea con matrice di transizione $\mathbf{P}$, vedi Teorema 19.9. Se $\mathbf{P}$ *è irriducibile*, allora $\frac{1}{z}\pi$ è l'unico vettore stocastico che è punto fisso di $P(x) = x\mathbf{P}$. Inoltre, per ogni $j \in S$ il numero $(\frac{1}{z}\pi_j)^{-1}$ è il tempo di ritorno in j per la catena $(X_n)_n$, vedi Corollario 21.5. Il teorema ergodico, (ii) del Teorema 21.3, assicura la convergenza

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k(x)) \;\to\; \frac{1}{z}\sum_{j\in S} f(j)\pi_j = E$$

per quasi ogni $x \in \Omega$.

Per simulare la catena è sufficiente avere un ampio numero di variabili aleatorie indipendenti che seguono la distribuzione uniforme in $[0,1]$, vedi Teorema 19.9, perció questa procedura, conosciuta come *catena di Markov Monte Carlo* puó condurre ad algoritmi numerici per il calcolo approssimato di E, vedi, per esempio, [21].

La simulazione di una catena di Markov con matrice di transizione $\mathbf{P}$ consiste nella transizione da una posizione $i \in S$ al passo n in una nuova posizione j al passo $(n+1)$ seguendo la densità di probabilità $(p_{ij})_j$. In questo caso

$$p_{ij} = \min\left\{1, \frac{\pi_j}{\pi_i}\right\} q_{ij}.$$

Si procede dunque nel modo seguente: supponiamo di essere nello stato i al passo n. Selezioniamo $j \in S$ seguendo la densità di massa $(q_{ij})_j$. Allora

- se $\pi_i \leq \pi_j$ ci spostiamo nello stato j, $X_{n+1} = j$;

◦ se $\pi_i > \pi_j$ effettuiamo un esperimento di Bernoulli con parametro $p = \pi_j/\pi_i$. Si pone $X_{n+1} = j$ se si ottiene un successo e $X_{n+1} = i$ altrimenti.

Due fatti sono cruciali per le applicazioni numeriche. Innanzitutto va verificato che la matrice $\mathbf{P}$ sia irriducibile, cosa che si può verificare per molte applicazioni del metodo. Il secondo punto cruciale è ovviamente il numero di passi necessari per ottenere una stima di E con l'accuratezza desiderata. Non ci occuperemo di queste questioni che coinvolgono stime sia statistiche che analitiche e che sono ancora un campo di ricerca aperto.

Invitiamo il lettore alla lettura di [21] e dell'articolo [9] in cui sono presentate ulteriori applicazioni, quali problemi di ottimizzazione, del metodo della catena di Markov Monte Carlo. Qui presentiamo solo il seguente risultato, vedi [5].

21.6 Teorema (Metropolis et al.). *Se π non è costante e se la matrice $\mathbf{P}$ definita nella (21.11) dell'algoritmo di Metropolis è irriducibile, allora $\mathbf{P}$ è regolare.*

Dimostrazione. Dimostriamo che esiste $h \in S$ tale che $p_{hh} > 0$. Sia $C := \{i \,|\, \pi_i = \max_j \pi_j\}$. Ovviamente, $C \neq \emptyset$ e, per ipotesi, $C \neq S$. Quindi, vedi Esercizio 18.18, esistono $h \in C$ e $k \notin C$ tale che $q_{hk} > 0$ e $\pi_h > \pi_k$. Se $i \neq j$, allora $p_{ij} \leq q_{ij}$, e per definizione di $\mathbf{P}$

$$\begin{aligned} p_{hh} &= 1 - \sum_{j \neq h} p_{hj} = 1 - \sum_{j \neq h,k} p_{hj} - p_{hk} \\ &\geq 1 - \sum_{j \neq h,k} q_{hj} - \frac{\pi_k}{\pi_h} q_{hk} = 1 - \sum_{j \neq h} q_{hj} + q_{hk}\Big(1 - \frac{\pi_k}{\pi_h}\Big) \\ &= q_{hh} + q_{hk}\Big(1 - \frac{\pi_k}{\pi_h}\Big) > 0. \end{aligned}$$

La tesi segue dall'Esercizio 17.16. □