

ALMA MATER STUDIORUM · UNIVERSITÀ DI
BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea in Matematica

IL TEOREMA ERGODICO
E
LE MONTE CARLO MARKOV
CHAINS

Relatore:
Chiar.mo Prof.
Stefano Pagliarani

Presentata da:
Alessandro Borgia

Anno Accademico 2020/21

Introduzione

Le catene di Markov devono il loro nome al matematico russo Andrej Andreevič Markov, il quale introdusse questa classe di processi stocastici a tempo discreto nei primi anni del 1900. Lo scopo originario dello studioso era mostrare che alcuni risultati di convergenza validi per successioni di variabili aleatorie indipendenti identicamente distribuite, quali la legge dei grandi numeri, potevano essere estesi a processi che ammettono una piccola dipendenza dal passato. Questa dipendenza rende possibili molti comportamenti diversi, che fanno delle catene di Markov un oggetto più interessante da studiare rispetto alle successioni di variabili aleatorie indipendenti identicamente distribuite.

Negli anni successivi queste catene si sono dimostrate utilissime in più settori e vantano svariate applicazioni in fisica, biologia, ingegneria elettronica e altri campi. La proprietà caratterizzante le catene di Markov è una sorta di perdita di memoria: lo stato attuale della catena è la massima informazione che possiamo avere per determinare lo stato successivo, mentre gli stati precedenti non influiscono.

Il primo capitolo della tesi si sofferma sulle catene di Markov discrete, seguendo l'impostazione di [B]. L'interesse principale è per le catene omogenee e irriducibili, che saranno classificate come ricorrenti positive, ricorrenti nulle o transitorie. Si definisce inoltre la distribuzione stazionaria e se ne dimostrano esistenza e unicità per le catene ricorrenti positive. Il capitolo si conclude col teorema ergodico, il quale ci fornisce un importante risultato di convergenza per le catene omogenee, irriducibili, ricorrenti positive.

Il secondo capitolo segue l'impostazione di [R]. Si inizia trattando brevemente il metodo Monte Carlo classico, fornendone un esempio con il metodo di *acceptance-rejection*. Dopo aver introdotto i concetti di periodicità e reversibilità di una catena di Markov, si presenta il metodo *Monte Carlo Markov Chains* per catene a valori in uno spazio finito e se ne fornisce un esempio: l'algoritmo di *Metropolis-Hastings*. L'elaborato si conclude con degli esempi di applicazione dell'algoritmo.

Indice

Introduzione	i
1 Il teorema ergodico	1
1.1 Catene di Markov discrete	1
1.2 Stati ricorrenti e transitori	4
1.3 Misura invariante	6
1.4 Criterio della distribuzione stazionaria	10
1.5 Teorema ergodico	11
2 Monte Carlo Markov Chains	15
2.1 Metodo Monte Carlo	15
2.2 Periodicità	17
2.3 Monte Carlo Markov Chains	19
2.4 L'algoritmo di Metropolis-Hastings	19
2.4.1 Gibbs Sampler	24
Richiami di teoria della probabilità	27
Bibliografia	35

Capitolo 1

Il teorema ergodico

1.1 Catene di Markov discrete

Definizione 1.1.1 (Catena di Markov discreta). Un processo stocastico a tempo discreto $\{X_n\}_{n \geq 0}$ a valori in uno spazio discreto E si dice una *catena di Markov* (discreta) se vale la seguente:

$$P(X_{n+1} = j \mid X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j \mid X_n = i_n),$$

per ogni $n \geq 0$ e per ogni scelta di $i_0, \dots, i_n, j \in E$ tale che siano ben definite le probabilità condizionate. Nel proseguo della trattazione si farà sempre riferimento a catene discrete.

Definizione 1.1.2 (HMC). Una catena di Markov $\{X_n\}_{n \geq 0}$ si dice *omogenea* (HMC) se $\forall i, j \in E, P(X_{n+1} = j \mid X_n = i)$ non dipende da n . In tal caso definiamo:

$$p_{ij} := P(X_{n+1} = j \mid X_n = i).$$

La matrice $P := \{p_{ij}\}_{i,j \in E}$ è detta *matrice di transizione* di $\{X_n\}_{n \geq 0}$.

Visto che i p_{ij} sono probabilità di eventi e dato che partendo da uno stato i mi muoverò sicuramente in un altro stato $k \in E$, osserviamo che valgono le seguenti:

$$p_{ij} \geq 0, \quad \sum_{k \in E} p_{ik} = 1.$$

Una generica matrice che rispetta queste due proprietà si dice essere una *matrice stocastica*.

Osservazione 1.1. Consideriamo una HMC $\{X_n\}_{n \geq 0}$, con matrice di transizione P . Si dice che la variabile aleatoria X_0 è lo *stato iniziale della catena*, inoltre la sua distribuzione v ,

$$v(i) = P(X_0 = i),$$

è detta *distribuzione iniziale*. Dalla formula di moltiplicazione abbiamo che

$$P(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) =$$

$$P(X_0 = i_0)P(X_1 = i_1 | X_0 = i_0) \cdots P(X_k = i_k | X_{k-1} = i_{k-1}, \dots, X_0 = i_0),$$

da cui per la proprietà di Markov omogenea e dalla definizione di matrice di transizione

$$P(X_0 = i_0, X_1 = i_1, \dots, X_k = i_k) = v(i_0)p_{i_0 i_1} \cdots p_{i_{k-1} i_k}.$$

Abbiamo dunque dimostrato la seguente:

Proposizione 1.1.1. *La distribuzione di una HMC dipende soltanto dalla sua distribuzione iniziale e dalla matrice di transizione.*

Osservazione 1.2. La distribuzione della catena al tempo n è il vettore v_n , dove

$$v_n(i) = P(X_n = i).$$

Dalla formula della probabilità totale abbiamo che $v_{n+1}(j) = \sum_{i \in E} v_n(i)p_{ij}$, ossia in forma vettoriale:

$$v_{n+1}^T = v_n^T P.$$

Iterando l'equazione otteniamo

$$v_n^T = v_0^T P^n.$$

$P^n := \{p_{ij}(n)\}_{i,j \in E}$ è la matrice di transizione a n passi, dato che in termini generali

$$p_{ij}(m) = P(X_{n+m} = j | X_n = i).$$

Osservazione 1.3. La proprietà di Markov si estende alla forma

$$\begin{aligned} P(X_{n+1} = j_1, \dots, X_{n+k} = j_k \mid X_n = i, \dots, X_0 = i_0) = \\ = P(X_{n+1} = j_1, \dots, X_{n+k} = j_k \mid X_n = i_n), \end{aligned}$$

per tutti gli $i_0, \dots, i_{n-1}, i, j_1, \dots, j_k$ per cui le probabilità condizionate sono ben definite. Scrivendo

$$A = \{X_{n+1} = j_1, \dots, X_{n+k} = j_k\}, \quad B = \{X_0 = i_0, \dots, X_{n-1} = i_{n-1}\},$$

troviamo che l'equazione precedente si scrive come

$$P(A \mid (X_n = i) \cap B) = P(A \mid X_n = i).$$

Moltiplicando ambo i membri per $P(B \mid X_n = i)$ e sfruttando la definizione di probabilità condizionata di eventi otteniamo che

$$P(A \cap B \mid X_n = i) = P(A \mid X_n = i)P(B \mid X_n = i).$$

A parole, il futuro al tempo n e il passato al tempo n , condizionati rispetto allo stato attuale $X_n = i$, sono indipendenti.

Osservazione 1.4 (Notazione). Utilizziamo la notazione $P_i(A)$ per indicare $P(A \mid X_0 = i)$ e E_i per l'attesa condizionata a $(X_0 = i)$. Inoltre se μ è una distribuzione di probabilità su E , allora indichiamo con $P_\mu(A) := \sum_{i \in E} \mu(i)P_i(A)$ la probabilità di A data μ come distribuzione iniziale.

Definizione 1.1.3 (Stati comunicanti). Sia $\{X_n\}_{n \geq 0}$ una HMC. Due stati $i, j \in E$ si dicono *comunicanti* se esistono $M, N \geq 0$ tali che $p_{ij}(M) > 0, p_{ji}(N) > 0$.

Osserviamo che uno stato comunica con se stesso dato che $p_{ii}(0) = 1$; più in generale osserviamo che la comunicazione è una relazione di equivalenza che genera una partizione di E in classi.

Definizione 1.1.4 (Irriducibilità). Una catena di Markov omogenea e la sua matrice di transizione si dicono *irriducibili* se ogni stato comunica con tutti gli altri o, equivalentemente, se c'è una sola classe di comunicazione.

Definizione 1.1.5 (Tempo di ritorno). Sia $i \in E$. Indichiamo il *tempo di ritorno* allo stato i con T_i , dove

$$T_i := \inf\{n \geq 1 : X_n = i\} \in \mathbb{N}_{\geq 1} \cup \infty.$$

Teorema 1.1.2 (Proprietà di Markov forte). Siano $\{X_n\}_{n \geq 0}$ una HMC e P la sua matrice di transizione. Sia $0 \in E$ uno stato qualsiasi e sia T_0 il tempo di ritorno in 0 . Valgono le seguenti:

- Il processo dopo T_0 è indipendente dal processo prima di T_0 ;
- Il processo dopo T_0 è una HMC con matrice di transizione P .

Dimostrazione. Vedi [B]. □

1.2 Stati ricorrenti e transitori

Definizione 1.2.1 (Stati ricorrenti e transitori). Uno stato $i \in E$ è detto *ricorrente* se il tempo di ritorno in i condizionato rispetto ad $X_0 = i$ è P-quasi certamente finito, ovvero se

$$P_i(T_i < \infty) = 1.$$

Uno stato si dice *ricorrente positivo* se

$$E_i[T_i] < \infty.$$

Uno stato ricorrente non positivo si dice *ricorrente nullo*. Uno stato non ricorrente si dice *transitorio*.

Teorema 1.2.1 (Teorema del ciclo rigenerativo). Sia $\{X_n\}_{n \geq 0}$ una HMC, $0 \in E$ uno stato ricorrente. Denotiamo con $\tau_0, \tau_1, \tau_2, \dots$ i tempi in cui si visita lo stato 0 , dove $\tau_0 = 0, \tau_1 = T_0$. Allora i pezzi di traiettoria

$$\{X_{\tau_k}, X_{\tau_k+1}, \dots, X_{\tau_{k+1}-1}\}, k \geq 0,$$

sono indipendenti e identicamente distribuiti. Questi pezzi di traiettoria si chiamano *cicli rigenerativi della catena tra le visite in 0* .

Dimostrazione. Segue direttamente dalla proprietà di Markov forte. \square

Definizione 1.2.2 (Matrice potenza). Data una matrice di transizione P , la *matrice potenza* G associata è definita come

$$G := \sum_{n \geq 0} P^n.$$

Dato che

$$g_{ij} = \sum_{n=0}^{\infty} p_{ij}(n) = \sum_{n=0}^{\infty} P_i(X_n = j) = \sum_{n=0}^{\infty} E_i [\mathbb{1}_{\{X_n=j\}}] = E_i \left[\sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=j\}} \right],$$

osserviamo che g_{ij} rappresenta il numero medio di visite allo stato j , condizionato al fatto che lo stato di partenza sia i .

Teorema 1.2.2 (Criterio della matrice potenza). *Uno stato $i \in E$ è ricorrente se e solo se*

$$\sum_{n=0}^{\infty} p_{ii}(n) = \infty.$$

Uno stato $i \in E$ è transitorio se e solo se

$$\sum_{n=0}^{\infty} p_{ii}(n) < \infty.$$

Dimostrazione. Vedi [B]. \square

Teorema 1.2.3 (La ricorrenza è una proprietà delle classi). *Due stati comunicanti sono entrambi transitori o entrambi ricorrenti.*

Dimostrazione. Per definizione, se due stati $i, j \in E$ sono comunicanti allora esistono $M, N \geq 0$ tali che $p_{ij}(M) > 0, p_{ji}(N) > 0$. Osserviamo che è possibile trovare un percorso che parte e arriva in i in $M + N + n$ passi: è sufficiente considerare il percorso che parte in i , arriva in j all'istante M , si trova di nuovo in j all'istante $M + n$ e infine torna in i all'istante $M + N + n$. Dato che questo è solo uno dei possibili percorsi, possiamo scrivere:

$$p_{ii}(M + N + n) \geq p_{ij}(M)p_{jj}(n)p_{ji}(N).$$

E con ragionamento analogo:

$$p_{jj}(M + N + n) \geq p_{ji}(N)p_{ii}(n)p_{ij}(M).$$

Definendo il numero strettamente positivo $\alpha := p_{ij}(M)p_{ji}(N)$, otteniamo le seguenti:

- $p_{jj}(M + N + n) \geq \alpha p_{ii}(n)$,
- $p_{ii}(M + N + n) \geq \alpha p_{jj}(n)$.

Da cui $\sum_{n \geq 0} p_{ii}(n)$ converge se e solo se converge $\sum_{n \geq 0} p_{jj}(n)$. Concludiamo per il criterio della matrice potenza. \square

È possibile dimostrare che anche la ricorrenza positiva è una proprietà delle classi (vedi [B]).

Considerando una catena irriducibile tutti gli stati saranno della stessa natura ed è dunque possibile classificare la catena stessa come ricorrente positiva, ricorrente nulla o transitoria.

Definizione 1.2.3. Una HMC irriducibile si dice *ricorrente positiva* se uno stato $i \in E$ è ricorrente positivo (e di conseguenza tutti gli altri). Allo stesso modo si definiscono catene *ricorrenti nulle* e *catene transitorie*.

1.3 Misura invariante

Definizione 1.3.1 (Misura invariante). Un vettore non nullo $\{x_i\}_{i \in E}$ è chiamato *misura invariante* per una matrice stocastica $P = \{p_{ij}\}_{i,j \in E}$ se

$$x_i \in [0, \infty),$$

$$x_i = \sum_{j \in E} x_j p_{ji}.$$

$\forall i \in E$. Scrivendo la seconda proprietà in forma vettoriale, $x^T P = x^T$.

Definizione 1.3.2 (Distribuzione stazionaria). Data una matrice stocastica P , una distribuzione di probabilità π si dice essere una *distribuzione stazionaria* per P se soddisfa

$$\pi^T = \pi^T P.$$

Osservazione 1.5. Iterando l'operazione si ha che $\pi^T = \pi^T P^n$ per ogni istante n e dunque una catena iniziata da una distribuzione stazionaria mantiene sempre la stessa distribuzione.

Teorema 1.3.1 (Forma canonica della misura invariante). *Sia $\{X_n\}_{n \geq 0}$ una HMC irriducibile e ricorsiva, P la sua matrice di transizione. Siano $0 \in E$ uno stato qualsiasi e T_0 il tempo di ritorno in 0. Definita*

$$x_i := E_0 \left[\sum_{n \geq 1} \mathbb{1}_{\{X_n=i\}} \mathbb{1}_{\{n \leq T_0\}} \right] \quad \forall i \in E, \quad (1)$$

x è una misura invariante per P e in particolare $x_i > 0 \forall i \in E$.

Dimostrazione. Definiamo:

$${}_0p_{0i}(n) := E_0 \left[\mathbb{1}_{\{X_n=i\}} \mathbb{1}_{\{n \leq T_0\}} \right] = P_0(X_1 \neq 0, \dots, X_{n-1} \neq 0, X_n = i),$$

che rappresenta la probabilità di visitare i al tempo n prima di ritornare in 0, partendo dallo stato 0. Dalla definizione di x si ha:

$$x_i = \sum_{n \geq 1} {}_0p_{0i}(n).$$

È evidente che ${}_0p_{0i}(1) = p_{0i}$, inoltre osserviamo che per ogni $n \geq 2$:

$${}_0p_{0i}(n) = \sum_{j \neq 0} {}_0p_{0j}(n-1)p_{ji},$$

da cui otteniamo:

$$\sum_{n \geq 2} {}_0p_{0i}(n) = \sum_{j \neq 0} \sum_{n \geq 1} {}_0p_{0j}(n)p_{ji} = \sum_{j \neq 0} x_j p_{ji}.$$

Sommando con quanto ottenuto per $n = 1$:

$$x_i = p_{0i} + \sum_{j \neq 0} x_j p_{ji} = \sum_{j \in E} x_j p_{ji},$$

dove l'ultima equazione si ottiene osservando che $x_0 = 1$.

Mostriamo adesso che $x_i > 0 \forall i \in E$. Iterando $x^T P = x^T$ troviamo $x^T P^n = x^T$, ovvero:

$$x_i = \sum_{j \in E} x_j p_{ji}(n) = p_{0i}(n) + \sum_{j \neq 0} x_j p_{ji}(n) \quad \forall n \geq 1.$$

Dato che la sommatoria è non negativa, se per qualche stato i si annullasse x_i si annullerebbe anche $p_{0i}(n)$ per ogni n , il che vorrebbe dire che 0 non comunica con i contraddicendo l'ipotesi di irriducibilità.

Resta da mostrare che $x_i < \infty$ per ogni $i \in E$, ma osservando che

$$1 = x_0 = \sum_{j \in E} x_j p_{j0}(n) \quad \forall n \geq 1,$$

se per qualche i fosse $x_i = \infty$ dovrebbe essere $p_{i0}(n) = 0$ per ogni $n \geq 1$ e di nuovo si contraddirebbe l'irriducibilità. \square

Teorema 1.3.2 (Unicità della misura invariante). *Data una matrice stocastica P irriducibile e ricorrente esiste un'unica misura invariante a meno di moltiplicazioni per una costante.*

Dimostrazione. Nella dimostrazione precedente, per dimostrare che $x_i > 0 \forall i \in E$, con x misura invariante, abbiamo utilizzato soltanto l'irriducibilità e dunque questo risultato resta valido per una generica HMC irriducibile. Allora definita y come la misura invariante è ben posta la seguente definizione:

$$q_{ji} := \frac{y_i}{y_j} p_{ij} \quad \forall i, j \in E.$$

$Q = \{q_{ji}\}_{i,j \in E}$ è una matrice stocastica, infatti

$$\sum_{i \in E} q_{ji} = \frac{1}{y_j} \sum_{i \in E} y_i p_{ij} = \frac{y_j}{y_j} = 1.$$

Mostriamo per induzione che gli elementi di Q^n sono $q_{ji}(n) = \frac{y_i}{y_j} p_{ij}(n)$. Supponiamolo vero per n e mostriamolo per $n+1$:

$$q_{ji}(n+1) = \sum_{k \in E} q_{jk} q_{ki}(n) = \sum_{k \in E} \frac{y_k}{y_j} p_{kj} \frac{y_i}{y_k} p_{ki}(n)$$

$$= \frac{y_i}{y_j} \sum_{k \in E} p_{ik}(n) p_{kj} = \frac{y_i}{y_j} p_{ij}(n+1).$$

Dunque vale per ogni n .

Osservando che $q_{ii}(n) = p_{ii}(n)$, segue direttamente dal criterio della matrice potenziale che Q è ricorrente, inoltre dalla definizione è facile osservare come Q sia irriducibile dato che lo è P .

Sia $g_{ji}(n)$ la probabilità (relativa alla catena governata da Q) di visitare per la prima volta lo stato i all'istante n partendo da j . Osserviamo che $g_{i0}(n+1) = \sum_{j \neq 0} q_{ij} g_{j0}(n)$; moltiplicando ambo i membri per y_i si ha

$$y_i g_{i0}(n+1) = \sum_{j \neq 0} (y_j g_{j0}(n)) p_{ji}.$$

Ricordando ${}_0p_{0i}(n+1) = \sum_{j \neq 0} {}_0p_{0j}(n) p_{ji}$ otteniamo che valgono le seguenti:

$$y_0 {}_0p_{0i}(n+1) = \sum_{j \neq 0} (y_0 {}_0p_{0j}(n)) p_{ji},$$

$$y_0 {}_0p_{0i}(1) = y_0 p_{0i},$$

$$y_i g_{i0}(1) = y_i q_{i0} = y_0 p_{0i}.$$

Dunque le sequenze $\{y_0 {}_0p_{0i}(n)\}$ e $\{y_i g_{i0}(n)\}$ soddisfano la stessa ricorrenza e hanno il primo termine uguale. Possiamo quindi concludere che

$${}_0p_{0i}(n) = \frac{y_i}{y_0} g_{i0}(n) \quad \forall n \geq 1.$$

Sommando per tutti gli n e sfruttando la ricorsività di Q otteniamo che

$$x_i = \frac{y_i}{y_0}.$$

□

Teorema 1.3.3. *Se una HMC irriducibile e ricorrente ammette una distribuzione stazionaria, allora questa è unica.*

Dimostrazione. Una distribuzione stazionaria è una misura invariante ed è quindi unica a meno di moltiplicazioni per una costante: l'unica costante moltiplicativa che la fa rimanere una distribuzione è 1. □

Teorema 1.3.4. *Una HMC irriducibile e ricorrente è ricorrente positiva se e solo se ogni sua misura invariante x è tale che:*

$$\sum_{i \in E} x_i < \infty.$$

Dimostrazione. Definendo x come la misura invariante canonica si ottiene che:

$$\sum_{i \in E} \sum_{n \geq 1} \mathbb{1}_{\{X_n=i\}} \mathbb{1}_{\{n \leq T_0\}} = \sum_{n \geq 1} \left(\sum_{i \in E} \mathbb{1}_{\{X_n=i\}} \right) \mathbb{1}_{\{n \leq T_0\}} = \sum_{n \geq 1} \mathbb{1}_{\{n \leq T_0\}} = T_0,$$

da cui, richiamando (1),

$$\sum_{i \in E} x_i = E_0 [T_0].$$

La tesi è una diretta conseguenza. \square

Osservazione 1.6. L'ipotesi di ricorrenza è necessaria: una HMC può essere irriducibile e ammettere una misura invariante senza essere ricorrente.

1.4 Criterio della distribuzione stazionaria

Teorema 1.4.1. *Sia $\{X_n\}_{n \geq 0}$ una catena di Markov omogenea e irriducibile, allora $\{X_n\}_{n \geq 0}$ è ricorrente positiva se e solo se ammette una distribuzione stazionaria. Inoltre, se la distribuzione π esiste, è unica e si ha che $\pi > 0$.*

Dimostrazione. Se la catena è ricorrente positiva allora consideriamo la misura invariante canonica x . Dato che $\sum_{i \in E} x_i < \infty$ possiamo ottenere una distribuzione stazionaria normalizzando x .

Sia invece π una distribuzione stazionaria. Ricordando che $\pi^T = \pi^T P^n \forall n$, abbiamo

$$\pi(i) = \sum_{j \in E} \pi(j) p_{ji}(n), \quad \forall i \in E.$$

Se la catena fosse transitoria, per ogni scelta di $i, j \in E$ sarebbe:

$$\lim_{n \rightarrow \infty} p_{ji}(n) = 0,$$

ma, dato che $p_{ji}(n)$ è limitato da 1, per la convergenza dominata otteniamo che

$$\pi(i) = \lim_{n \rightarrow \infty} \sum_{j \in E} \pi(j) p_{ji}(n) = \sum_{j \in E} \pi(j) \left(\lim_{n \rightarrow \infty} p_{ji}(n) \right) = 0.$$

Dunque se la catena fosse transitoria π non sarebbe una distribuzione stazionaria (in particolare $\sum_{i \in E} \pi(i) \neq 1$), per cui la catena deve essere ricorrente. Allora per il Teorema 1.3.3 la distribuzione stazionaria è unica ed è ottenuta normalizzando la misura invariante canonica, per cui $\pi > 0$. Inoltre, dato che $\sum_{i \in E} \pi(i)$ è una quantità finita (vale 1 per definizione), allora per il Teorema 1.3.4 la catena è ricorrente positiva. \square

Teorema 1.4.2. *Una HMC irriducibile a valori in uno spazio finito è ricorrente positiva.*

Dimostrazione. Proviamo innanzitutto la ricorrenza. Se per assurdo non fosse ricorrente, ogni stato sarebbe visitato un numero finito di volte ed esisterebbe un tempo finito dopo il quale nessuno stato potrebbe essere visitato. Sia x la misura invariante canonica, dato che E è finito $\sum_{i \in E} x_i < \infty$ e dunque normalizzando x troviamo la distribuzione stazionaria. Concludiamo per il criterio della distribuzione stazionaria. \square

1.5 Teorema ergodico

Proposizione 1.5.1. *Sia $\{X_n\}_{n \geq 0}$ una HMC irriducibile e ricorrente. Sia inoltre x la misura invariante canonica associata. Consideriamo uno stato $0 \in E$ e una funzione $f : E \rightarrow \mathbb{R}$ tale che*

$$\sum_{i \in E} |f(i)| x_i < \infty.$$

Definita

$$v(n) := \sum_{k=1}^n \mathbb{1}_{\{X_k=0\}},$$

per una qualsiasi distribuzione iniziale μ vale :

$$\lim_{N \rightarrow \infty} \frac{1}{v(N)} \sum_{k=1}^N f(X_k) = \sum_{i \in E} f(i) x_i \quad P_\mu - q.c..$$

Dimostrazione. Siano $\tau_1, \tau_2, \tau_3, \dots$ i tempi di ritorno allo stato $0 \in E$ con $\tau_1 = T_0$. Definiamo:

$$U_p := \sum_{n=\tau_p+1}^{\tau_{p+1}} f(X_n).$$

Osserviamo che, per il teorema del ciclo rigenerativo, le $\{U_p\}_{p \geq 1}$ sono una successione di variabili aleatorie i.i.d..

Assumendo $f \geq 0$ e usando la proprietà di Markov forte

$$E[U_1] = E \left[\sum_{n=\tau_1+1}^{\tau_2} f(X_n) \mid X_{\tau_1} = 0 \right] = E_0 \left[\sum_{n=1}^{T_0} f(X_n) \right],$$

proseguendo:

$$\begin{aligned} E_0 \left[\sum_{n=1}^{T_0} f(X_n) \right] &= E_0 \left[\sum_{n=1}^{T_0} \sum_{i \in E} f(i) \mathbb{1}_{\{X_n=i\}} \right] = \\ &= \sum_{i \in E} f(i) E_0 \left[\sum_{n=1}^{T_0} \mathbb{1}_{\{X_n=i\}} \right] = \sum_{i \in E} f(i) x_i. \end{aligned}$$

In definitiva,

$$E[U_1] = \sum_{i \in E} f(i) x_i.$$

Per ipotesi questa somma è una quantità finita e applicando la legge dei grandi numeri forte otteniamo che

$$\sum_{i \in E} f(i) x_i = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{p=1}^n U_p = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=T_0+1}^{\tau_{n+1}} f(X_k).$$

Dato che vale

$$\tau_{v(n)} \leq n < \tau_{v(n)+1},$$

otteniamo:

$$\frac{\sum_{k=1}^{\tau_{v(n)}} f(X_k)}{v(n)} \leq \frac{\sum_{k=1}^n f(X_k)}{v(n)} \leq \frac{\sum_{k=1}^{\tau_{v(n)+1}} f(X_k)}{v(n)}.$$

La ricorrenza della catena ci dice che $v(n) \xrightarrow{n \rightarrow \infty} \infty$, da cui entrambi i termini agli estremi della catena di disequazioni tendono a $\sum_{i \in E} f(i) x_i$, da cui si ottiene:

$$\lim_{N \rightarrow \infty} \frac{1}{v(N)} \sum_{k=1}^N f(X_k) = \sum_{i \in E} f(i) x_i \quad P_\mu \text{ q.c..}$$

In caso di funzione di segno arbitrario si arriva alla tesi considerando separatamente f^+ e f^- e seguendo il medesimo procedimento.

□

Teorema 1.5.2. *Sia $\{X_n\}_{n \geq 0}$ una catena di Markov irriducibile e ricorrente positiva. Sia π la rispettiva distribuzione stazionaria e consideriamo una $f : E \rightarrow \mathbb{R}$ tale che*

$$\sum_{i \in E} |f(i)|\pi(i) < \infty.$$

Allora, per ogni distribuzione iniziale μ , vale la seguente:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N f(X_k) = \sum_{i \in E} f(i)\pi(i) \quad P_\mu - q.c..$$

Dimostrazione. Possiamo applicare la Proposizione 1.5.1 utilizzando la funzione costante $f \equiv 1$. Tutte le ipotesi sono soddisfatte in quanto per la ricorrenza positiva si ha che $\sum_{i \in E} x_i < \infty$, allora otteniamo che

$$\lim_{N \rightarrow \infty} \frac{N}{v(N)} = \sum_{i \in E} x_i \quad P_\mu - q.c..$$

Dato che la distribuzione stazionaria è proporzionale alla misura invariante canonica, si ha che f soddisfa $\sum_{i \in E} |f(i)|x_i < \infty$ e dunque

$$\lim_{N \rightarrow \infty} \frac{1}{v(N)} \sum_{k=1}^N f(X_k) = \sum_{i \in E} f(i)x_i \quad P_\mu - q.c..$$

Combinando le due equazioni si ha che

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N f(X_k) = \lim_{N \rightarrow \infty} \frac{v(N)}{N} \frac{1}{v(N)} \sum_{k=1}^N f(X_k) = \frac{\sum_{i \in E} f(i)x_i}{\sum_{j \in E} x_j} \quad P_\mu - q.c.,$$

da cui si ottiene la tesi visto che π si ottiene normalizzando x .

□

Capitolo 2

Monte Carlo Markov Chains

2.1 Metodo Monte Carlo

Sia Z una variabile aleatoria k -dimensionale con densità $f(x)$. Sia inoltre $\Phi : \mathbb{R}^k \rightarrow \mathbb{R}$ misurabile e tale $E[|\Phi(Z)|] < \infty$. Supponendo di voler calcolare $E[\Phi(Z)]$, la formula

$$E[\Phi(Z)] = \int_{\mathbb{R}^k} \Phi(x) f(x) dx$$

è utilizzabile se sappiamo calcolare analiticamente l'integrale; una alternativa è rappresentata dal *metodo Monte Carlo*.

Il metodo Monte Carlo consiste nel generare una sequenza di variabili aleatorie $\{Z_n\}_{n \geq 0}$ i.i.d. con la stessa distribuzione di Z . Sfruttando la legge dei grandi numeri forte, per cui

$$E[\Phi(Z)] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Phi(Z_i) \quad P - q.c.,$$

possiamo ottenere una stima del valore atteso basata su un campione finito (Z_1, \dots, Z_n) .

Osservazione 2.1. Per utilizzare il metodo Monte Carlo è necessario essere in grado di generare variabili aleatorie con la stessa distribuzione di Z . Uno dei metodi più utilizzati è il metodo *acceptance-rejection* (A-R).

Teorema 2.1.1 (Metodo A-R). *Sia $\{X_n\}_{n \geq 0}$ una successione di variabili aleatorie i.i.d. con legge assolutamente continua e densità $g(x)$, dove g è tale che*

$$\frac{f(x)}{g(x)} \leq c,$$

per una qualche costante c . Sia inoltre $\{U_n\}_{n \geq 0}$ una successione di v.a. i.i.d. con distribuzione $Unif_{[0,1]}$, indipendente da $\{X_n\}_{n \geq 0}$. Definiamo:

$$\tau_0 := 0,$$

$$\tau_n := \min \left\{ n \geq \tau_{n-1} \mid U_n \leq \frac{f(X_n)}{cg(X_n)} \right\}.$$

Allora ponendo $Y_n = X_{\tau_n}$, $\{Y_n\}_{n \geq 0}$ è una successione di variabili aleatorie i.i.d. con distribuzione assolutamente continua e densità $f(x)$.

Dimostrazione. Una modifica della dimostrazione del teorema del ciclo rigenerativo permette di verificare che $\{Y_n\}_{n \geq 0}$ è una successione di variabili aleatorie i.i.d., con la distribuzione di X_1 condizionata a $\left(U_1 \leq \frac{f(X_1)}{cg(X_1)} \right)$. Perciò per ogni $A \in \mathcal{B}$

$$P(Y_1 \in A) = P \left((X_1 \in A) \mid \left(U_1 \leq \frac{f(X_1)}{cg(X_1)} \right) \right) = \frac{P \left((X_1 \in A) \cap \left(U_1 \leq \frac{f(X_1)}{cg(X_1)} \right) \right)}{P \left(U_1 \leq \frac{f(X_1)}{cg(X_1)} \right)}.$$

Studiamo il denominatore. Per la formula della probabilità totale sarà:

$$P \left(U_1 \leq \frac{f(X_1)}{cg(X_1)} \right) = E \left[\mathbb{1}_{\left(U_1 \leq \frac{f(X_1)}{cg(X_1)} \right)} \right] = E \left[E \left[\mathbb{1}_{\left(U_1 \leq \frac{f(X_1)}{cg(X_1)} \right)} \mid X_1 \right] \right].$$

Dato che U_1 e X_1 sono indipendenti possiamo applicare il lemma di Freezing.

Ricordando che U_1 ha distribuzione $Unif_{[0,1]}$ si ha

$$E \left[E \left[\mathbb{1}_{\left(U_1 \leq \frac{f(X_1)}{cg(X_1)} \right)} \mid X_1 \right] \right] = E \left[\frac{f(X_1)}{cg(X_1)} \right].$$

Siccome X_1 ha densità $g(x)$

$$E \left[\frac{f(X_1)}{cg(X_1)} \right] = \int_{\mathbb{R}} \frac{f(x)}{cg(x)} g(x) dx = \frac{1}{c}.$$

Allora otteniamo che

$$P(Y_1 \in A) = cP \left((X_1 \in A) \cap \left(U_1 \leq \frac{f(X_1)}{cg(X_1)} \right) \right).$$

Utilizzando la formula della probabilità totale,

$$P \left((X_1 \in A) \cap \left(U_1 \leq \frac{f(X_1)}{cg(X_1)} \right) \right) = E \left[E \left[\mathbb{1}_{(X_1 \in A)} \mathbb{1}_{(U_1 \leq \frac{f(X_1)}{cg(X_1)})} \mid X_1 \right] \right],$$

da cui, dato che $\mathbb{1}_{(X_1 \in A)} \in \sigma(X_1)$ ed è limitata,

$$\begin{aligned} P(Y_1 \in A) &= cE \left[\mathbb{1}_{(X_1 \in A)} P \left(U_1 \leq \frac{f(X_1)}{cg(X_1)} \mid X_1 \right) \right] = cE \left[\mathbb{1}_{(X_1 \in A)} \frac{f(X_1)}{cg(X_1)} \right] = \\ &= c \int_A \frac{f(x)}{cg(x)} g(x) dx = \int_A f(x) dx. \end{aligned}$$

Abbiamo dunque ottenuto che

$$P(Y_1 \in A) = \int_A f(x) dx,$$

ossia Y_1 ha distribuzione assolutamente continua con densità f . \square

Osservazione 2.2. Per utilizzare il metodo A-R è indispensabile saper simulare una successione di variabili aleatorie i.i.d. $\{X_n\}_{n \geq 0}$ con distribuzione assolutamente continua e densità $g(x)$.

2.2 Periodicità

Definizione 2.2.1 (Stati periodici e aperiodici). Sia $\{X_n\}_{n \geq 0}$ una catena di Markov omogenea, definito il *periodo* di uno stato $i \in E$ come

$$d_i := \gcd\{n \geq 1 \mid p_{ii}(n) > 0\},$$

diciamo che i è *periodico* se $d > 1$, ossia se partendo da i ci torno con probabilità non nulla solo in tempi t multipli di $d > 1$. Se $d = 1$ lo stato si dice *aperiodico*.

Teorema 2.2.1 (La periodicità è una proprietà delle classi). *Due stati comunicanti hanno lo stesso periodo.*

Dimostrazione. Per definizione esistono $M, N > 0$ tali che $p_{ij}(M) > 0$ e $p_{ji}(N) > 0$. Osservato che

$$p_{ii}(M + kn + N) \geq p_{ij}(M)p_{jj}(k)^n p_{ji}(N),$$

si ha che se $p_{jj}(k) > 0$ allora $p_{ii}(M + kn + N) > 0$ per ogni $n \geq 1$. Dunque d_i divide $M + kn + N$ per ogni n , da cui in particolare d_i divide k . Dato che vale per tutti i k tali che $p_{jj}(k) > 0$, otteniamo che d_i divide d_j . Per simmetria d_j divide d_i , da cui l'uguaglianza. \square

Definizione 2.2.2 (Catene periodiche e aperiodiche). Una HMC irriducibile si dice *aperiodica* se un suo stato è aperiodico (e di conseguenza tutti gli altri). Altrimenti si dice *periodica*.

Definizione 2.2.3 (Catena ergodica). Una HMC irriducibile, ricorrente positiva e aperiodica si dice *ergodica*.

Osservazione 2.3. Si può mostrare che se la catena è aperiodica esiste la *probabilità limite* (vedi [B]) definita come

$$l_j := \lim_{n \rightarrow \infty} P(X_n = j).$$

Se lavoriamo in uno spazio finito, osservando che

$$P(X_{n+1} = j) = \sum_{i \in E} P(X_{n+1} = j \mid X_n = i)P(X_n = i) = \sum_{i \in E} p_{ij}P(X_n = i)$$

e

$$\sum_{i \in E} P(X_n = i) = 1,$$

mandando n a infinito troviamo che l è la distribuzione stazionaria. Osserviamo che occorre lavorare su uno spazio finito per poter portare il limite dentro la sommatoria, il risultato resta comunque valido per spazi discreti (vedi [B]).

Il vantaggio di costruire una catena aperiodica su uno spazio finito è dunque quello di non dover trovare direttamente la distribuzione stazionaria, in quanto la distribuzione di X_n vi convergerà al crescere di n .

2.3 Monte Carlo Markov Chains

Sia X una variabile aleatoria a valori in uno spazio finito E , con distribuzione π nota a meno di una costante e supponiamo di voler generare una variabile aleatoria con distribuzione π . Come esempio si può considerare di non conoscere la cardinalità di E e di voler generare valori con distribuzione uniforme.

Il metodo *Monte Carlo Markov Chains* (MCMC) consiste nel costruire una HMC irriducibile aperiodica $\{X_n\}_{n \geq 0}$ a valori in E , con distribuzione stazionaria π . Osserviamo che la distribuzione stazionaria esiste in quanto una HMC irriducibile a valori in uno spazio finito è ricorrente positiva (Teorema 1.4.2). Dato che la catena è ergodica avremo che per ogni stato $i \in E$ e per ogni distribuzione iniziale μ

$$\lim_{n \rightarrow \infty} P_\mu(X_n = i) = \pi(i)$$

e per il teorema ergodico

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \Phi(X_n) = E[\Phi(X)],$$

dove $\Phi : E \rightarrow R$ misurabile. Possiamo dunque generare una variabile aleatoria con distribuzione che tende a π e approssimare il valore atteso di $\Phi(X)$ basandoci su un campione finito (X_1, \dots, X_N) .

A questo punto bisogna costruire un algoritmo MCMC (Monte Carlo Markov Chains) che ci permetta di generare una HMC con le caratteristiche sopra elencate.

2.4 L'algoritmo di Metropolis-Hastings

Sia E finito, $\pi = \bar{\pi}/B$ una distribuzione di probabilità, dove è dato sapere la classe di distribuzioni alla quale $\bar{\pi}$ appartiene e B è una costante non nota. L'obiettivo è costruire una HMC irriducibile, con distribuzione stazionaria π . Supponiamo ad esempio di non conoscere la cardinalità di E e di voler

generare valori con distribuzione uniforme. In questo caso sarà $\bar{\pi} = 1$ e $B = |E|$. Ovviamente esistono infinite possibili scelte di $\bar{\pi}$ e di B e in generale diciamo che π è nota a meno di una costante (in questo caso, la cardinalità dell'insieme).

Chiamiamo $\{X_n\}_{n \geq 0}$ la catena, P la rispettiva matrice di transizione e procediamo per iterazione seguendo l'algoritmo di *Metropolis-Hastings*. Scelta arbitrariamente una distribuzione iniziale, all'istante $n + 1$ l'algoritmo propone uno stato successivo tramite una matrice stocastica irriducibile $Q = \{q_{ij}\}_{i,j \in E}$. In altre parole, se ci troviamo nello stato i , lo stato j viene proposto con probabilità q_{ij} . A questo punto j viene accettato con probabilità α_{ij} e viene rifiutato con probabilità $1 - \alpha_{ij}$. In caso di rifiuto, si pone $X_{n+1} = i$. Definiamo α_{ij} , $\forall i, j \in E$, ponendo:

$$\alpha_{ij} := \min \left(1, \frac{\bar{\pi}(j)q_{ji}}{\bar{\pi}(i)q_{ij}} \right).$$

Dunque, se $X_n = i$ e lo stato proposto è j , abbiamo:

$$X_{n+1} = \begin{cases} j & \text{con probabilità } \alpha_{ij}, \\ i & \text{altrimenti.} \end{cases}$$

Per passare da i a $j \neq i$ occorre che j sia prima proposto e poi accettato. Inoltre abbiamo che $X_{n+1} = i$ sia se lo stato proposto viene rifiutato, sia se è i stesso. Siamo allora in grado di calcolare gli elementi di $P = \{p_{ij}\}_{i,j \in E}$:

$$p_{ij} = \begin{cases} q_{ij}\alpha_{ij} & \text{se } j \neq i, \\ q_{ii} + \sum_{k \neq i} q_{ik}(1 - \alpha_{ik}) & \text{se } j = i. \end{cases} \quad (2)$$

Per dimostrare il funzionamento dell'algoritmo di Metropolis-Hastings introduciamo il concetto di *reversibilità* di una catena di Markov.

Definizione 2.4.1 (Catena a tempo invertito). Sia $\{X_n\}_{n \geq 0}$ una HMC ricorrente positiva, con matrice di transizione P . Ricordiamo che $\{X_n\}_{n \geq 0}$ ammette una distribuzione stazionaria π per il criterio della distribuzione stazionaria e π è tale che

$$\pi(i) > 0$$

per ogni stato $i \in E$. Definiamo la *matrice a tempo invertito* Q , indicizzata su E , tramite la seguente:

$$\pi(i)q_{ij} = \pi(j)p_{ji}.$$

La matrice è una matrice stocastica dato che

$$\sum_{j \in E} q_{ij} = \sum_{j \in E} \frac{\pi(j)}{\pi(i)} p_{ji} = \frac{1}{\pi(i)} \sum_{j \in E} \pi(j) p_{ji} = \frac{\pi(i)}{\pi(i)} = 1.$$

Osservazione 2.4. Supponiamo che la distribuzione iniziale di $\{X_n\}_{n \geq 0}$ sia π , dalla definizione di distribuzione stazionaria abbiamo che per ogni $n \geq 0, i \in E$,

$$P(X_n = i) = \pi(i).$$

Dalla formula di Bayes,

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

ossia

$$P(X_n = j \mid X_{n+1} = i) = q_{ji}.$$

Osserviamo che Q è la matrice di transizione della catena iniziale ma con il tempo "invertito".

Teorema 2.4.1 (Test di reversibilità). *Sia P una matrice stocastica indicizzata su un insieme numerabile E e sia π una misura di probabilità su E . Sia inoltre Q una matrice stocastica indicizzata su E tale che $\forall i, j \in E$*

$$\pi(i)q_{ij} = \pi(j)p_{ji}.$$

Allora π è la distribuzione stazionaria di P .

Dimostrazione. Fissiamo $i \in E$, considerando che $\pi(i)q_{ij} = \pi(j)p_{ji}$ e sommando rispetto a j otteniamo

$$\sum_{j \in E} \pi(i)q_{ij} = \sum_{j \in E} \pi(j)p_{ji}.$$

Ma dato che

$$\sum_{j \in E} \pi(i) q_{ij} = \pi(i) \sum_{j \in E} q_{ij} = \pi(i),$$

concludiamo che

$$\pi(i) = \sum_{j \in E} \pi(j) p_{ji}.$$

□

Definizione 2.4.2 (Catena reversibile). Consideriamo una HMC $\{X_n\}_{n \geq 0}$ ricorrente positiva, con matrice di transizione P e matrice a tempo invertito Q . Se la catena ha come distribuzione iniziale la distribuzione stazionaria π , diciamo che è *reversibile* se $P = Q$.

Osservazione 2.5. Dato che la distribuzione di una catena di Markov è totalmente determinata dalla sua distribuzione iniziale e dalla matrice di transizione, una catena reversibile e la rispettiva catena a tempo invertito sono statisticamente la stessa cosa.

Teorema 2.4.2. *La matrice stocastica P , definita come in (2), è irriducibile, reversibile e ha π come distribuzione stazionaria.*

Dimostrazione. L'irriducibilità segue direttamente dall'irriducibilità di Q . Per quanto riguarda la reversibilità è sufficiente mostrare che

$$\pi(i) p_{ij} = \pi(j) p_{ji}$$

$\forall i, j \in E$.

Dato che l'equazione è ovvia per $j = i$, supponiamo $j \neq i$ e mostriamo che

$$\pi(i) q_{ij} \alpha_{ij} = \pi(j) q_{ji} \alpha_{ji}.$$

Osserviamo che

$$\min \left(1, \frac{\pi(j) q_{ji}}{\pi(i) q_{ij}} \right) = \min \left(1, \frac{\bar{\pi}(j) q_{ji}}{\bar{\pi}(i) q_{ij}} \right) = \alpha_{ij}$$

e supponiamo che sia $\alpha_{ij} = \frac{\pi(j)q_{ji}}{\pi(i)q_{ij}}$. Allora sarà $\alpha_{ji} = 1$ e sostituendo otteniamo

$$\pi(i)q_{ij} \frac{\pi(j)q_{ji}}{\pi(i)q_{ij}} = \pi(j)q_{ji},$$

che è ovviamente verificata. \square

Osservazione 2.6. Osserviamo che in generale la catena costruita dall'algoritmo di Metropolis-Hastings non è aperiodica e quindi non garantisce l'esistenza di una probabilità limite. Per aggirare questo ostacolo possiamo inserire una piccola restrizione nella scelta di Q che garantisca l'esistenza di qualche stato i per cui $p_{ii} > 0$, ad esempio richiedere che la matrice Q non abbia diagonale identicamente nulla. A quel punto qualsiasi sia la distribuzione iniziale scelta, la catena costruita sarà aperiodica e la distribuzione di X_n convergerà alla distribuzione stazionaria π al crescere di n .

Osservazione 2.7. Dato che i primi valori sono condizionati dalla scelta della distribuzione iniziale, si utilizza una tecnica chiamata *burning* che consiste nello scartare i primi valori degli N campionati. Inoltre se si vuole soltanto generare uno o più valori con distribuzione π , è sufficiente non campionare l'intera sequenza bensì solo alcuni valori. La tecnica in questione è detta *thinning*.

Esempio. Sia E l'insieme delle n -ple (x_1, \dots, x_n) ottenibili tramite una permutazione dei numeri $(1, \dots, n)$ e tali che $\sum_{j=1}^n jx_j > a$ per una certa costante $a > 0$ data. Supponiamo di voler generare un elemento da E con distribuzione uniforme $\pi(i) = \frac{1}{|E|}$, senza necessariamente dover calcolare la cardinalità di E . Per farlo applichiamo l'algoritmo di Metropolis-Hastings.

Prima di tutto definiamo il concetto di vicinanza in E : diciamo che due n -ple in E sono vicine se sono ottenibili l'una dall'altra applicando una permutazione che scambi solo due elementi. Ad esempio $(1, 2, 3, 4)$ e $(1, 2, 4, 3)$ sono vicine. Indichiamo con $N(i)$ l'insieme degli elementi vicini a i e immaginiamo di costruire un grafo i cui vertici sono gli elementi di E , i quali sono collegati da un arco solo se vicini. La probabilità di transizione è così

definita:

$$q_{ij} = \begin{cases} \frac{1}{|N(i)|} & \text{se } j \in N(i), \\ 0 & \text{altrimenti.} \end{cases}$$

In altre parole, lo stato da proporre è scelto uniformemente tra le permutazioni vicine. Dato che la distribuzione target è uniforme abbiamo che $\pi(i) = \pi(j)$ e dunque

$$\alpha_{ij} = \min\left(\frac{|N(i)|}{|N(j)|}, 1\right).$$

In pratica se lo stato j proposto ha meno n -ple vicine rispetto ad i , allora j viene accettato; altrimenti si genera in modo uniformemente randomico un numero U nell'intervallo $(0, 1)$ e si accetta j se $U < \frac{|N(i)|}{|N(j)|}$, se ciò non si verifica lo stato successivo è i (j viene rifiutato). Dunque $p_{ii} > 0$ e la catena che si costruisce con l'algoritmo di Metropolis-Hastings sarà tale che la distribuzione di X_n converge a π . Siamo quindi in grado di generare elementi di E con distribuzione uniforme, anche senza calcolare la cardinalità dell'insieme.

2.4.1 Gibbs Sampler

Il *Gibbs Sampler* è una versione dell'algoritmo di Metropolis-Hastings. Sia $\mathbf{X} = (X_1, \dots, X_n)$ un vettore aleatorio discreto con distribuzione di probabilità $\pi(\mathbf{x})$ nota a meno di un fattore moltiplicativo e supponiamo di voler generare un vettore randomico con la distribuzione di \mathbf{X} , ossia vogliamo generare un vettore con distribuzione di probabilità

$$\pi(\mathbf{x}) = Cg(\mathbf{x}),$$

dove C non è nota. Per utilizzare il Gibbs Sampler assumiamo di saper generare una variabile aleatoria X con distribuzione

$$P(X = x) = P(X_i = x \mid X_j = x_j, j \neq i).$$

Utilizziamo l'algoritmo di Metropolis-Hastings su una catena di Markov con stati $\mathbf{x} = (x_1, \dots, x_n)$ e con probabilità di transizione definita come segue.

Se lo stato presente è \mathbf{x} , si sceglie uniformemente un indice i da 1 a n e si genera una variabile aleatoria X con distribuzione

$$P(X = x) = P(X_i = x \mid X_j = x_j, j \neq i).$$

Se $X = x$ lo stato proposto è $\mathbf{y} = (x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)$, ovvero la matrice Q è

$$q_{\mathbf{xy}} = \frac{1}{n} P(X_i = x \mid X_j = x_j, j \neq i) = \frac{\pi(\mathbf{y})}{nP(X_j = x_j, j \neq i)}.$$

Dall'algoritmo di Metropolis-Hastings abbiamo che \mathbf{y} viene accettato con probabilità

$$\alpha_{\mathbf{xy}} = \min \left(\frac{\pi(\mathbf{y})q_{\mathbf{yx}}}{\pi(\mathbf{x})q_{\mathbf{xy}}}, 1 \right) = \min \left(\frac{\pi(\mathbf{y})\pi(\mathbf{x})}{\pi(\mathbf{x})\pi(\mathbf{y})}, 1 \right) = 1,$$

ossia viene sempre accettato.

Appendice

Richiami di teoria della probabilità

In seguito vengono riportati risultati classici di teoria della probabilità. Per ulteriori approfondimenti si rimanda alla letteratura in materia; in particolare i seguenti risultati sono tratti da [P].

Definizione 0.1 (Spazio di probabilità). Uno *spazio di probabilità* è una terna (Ω, \mathcal{F}, P) dove:

- Ω è un insieme non vuoto;
- \mathcal{F} è una σ -algebra su Ω ;
- P è una misura tale che $P(\Omega) = 1$.

P si dice *misura di probabilità*.

Definizione 0.2 (Probabilità condizionata). Sia (Ω, \mathcal{F}, P) un spazio di probabilità, $B \in \mathcal{F}$ tale che $P(B) > 0$. Definiamo *la probabilità di A condizionata a B* come

$$P(A | B) := \frac{P(A \cap B)}{P(B)}.$$

Proposizione 0.1 (Formula della probabilità totale). Sia $\{B_i\}_{i \in I}$ una partizione discreta di Ω , con $P(B_i) > 0$ per ogni $i \in I$. Per ogni $A \in \mathcal{F}$ vale

$$P(A) = \sum_{i \in I} P(A | B_i)P(B_i).$$

Proposizione 0.2 (Formula di moltiplicazione). Siano $A_1, \dots, A_n \in \mathcal{F}$ tali che $P(A_1 \cap \dots \cap A_{n-1}) > 0$. Si ha

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2 | A_1) \dots P(A_n | A_1 \cap \dots \cap A_{n-1}).$$

Proposizione 0.3 (Formula di Bayes). Siano $A, B \in \mathcal{F}$ tali che valga $P(A), P(B) > 0$. Allora

$$P(B | A) = \frac{P(A | B)P(B)}{P(A)}.$$

Definizione 0.3 (Indipendenza di eventi). In uno spazio di probabilità (Ω, \mathcal{F}, P) , diciamo che due eventi A e B sono *indipendenti* in P se

$$P(A \cap B) = P(A)P(B).$$

Definizione 0.4 (Distribuzione). Sia \mathcal{B}_ρ la σ -algebra di Borel su uno spazio metrico (\mathbb{M}, ρ) . Una *distribuzione* è una misura di probabilità su $(\mathbb{M}, \mathcal{B}_\rho)$.

Definizione 0.5 (Delta di Dirac). Fissato $x_0 \in \mathbb{R}^d$, definiamo la distribuzione *delta di Dirac centrata in x_0* come

$$\delta_{x_0}(H) := \begin{cases} 1 & \text{se } x_0 \in H, \\ 0 & \text{altrimenti,} \end{cases}$$

$\forall H \in \mathcal{B}_d$.

Definizione 0.6 (Distribuzione discreta). Sia (x_n) una successione di punti distinti di \mathbb{R}^d e sia (p_n) una successione di numeri reali tali che

$$\sum_{n=1}^{\infty} p_n = 1, \quad p_n \geq 0, \quad n \in \mathbb{N}.$$

Una *distribuzione discreta* è una distribuzione della forma

$$\mu(H) := \sum_{n=1}^{\infty} p_n \delta_{x_n}(H), \quad H \in \mathcal{B}_d.$$

Definizione 0.7 (Distribuzione assolutamente continua). Una *densità* γ è una funzione \mathcal{B}_d -misurabile, $\gamma : \mathbb{R}^d \rightarrow [0, +\infty[$, tale che

$$\int_{\mathbb{R}^d} \gamma(x) dx = 1.$$

Si verifica facilmente che

$$\mu(H) := \int_H \gamma(x) dx, \quad H \in \mathcal{B}_d,$$

è una distribuzione. Una distribuzione si dice *assolutamente continua su* \mathbb{R}^d se esiste una densità γ per cui vale la scrittura precedente.

Definizione 0.8 (Variabile aleatoria). Siano $d \in \mathbb{N}$ fissato, (Ω, \mathcal{F}, P) uno spazio di probabilità. Data una funzione $X : \Omega \rightarrow \mathbb{R}^d$, definiamo per ogni $H \in \mathcal{B}_d$

$$(X \in H) := \{\omega \in \Omega \mid X(\omega) \in H\} = X^{-1}(H).$$

Diciamo che X è una *variabile aleatoria* se

$$(X \in H) \in \mathcal{F}, \quad \forall H \in \mathcal{B}_d.$$

Definizione 0.9. Definendo

$$\sigma(X) := \{(X \in H) \mid H \in \mathcal{B}_d\},$$

$\sigma(X)$ è una σ -algebra, detta *σ -algebra generata da* X .

Definizione 0.10 (Distribuzione di una variabile aleatoria). Data una variabile aleatoria X definita su (Ω, \mathcal{F}, P) , si chiama *distribuzione* (o *legge*) di X la distribuzione

$$\mu_X(H) := P(X \in H), \quad H \in \mathcal{B}_d.$$

Definizione 0.11 (Indipendenza di variabili aleatorie). Diciamo che le variabili aleatorie X_1, \dots, X_n definite su (Ω, \mathcal{F}, P) sono *indipendenti in* P se vale

$$P\left(\bigcap_{i=1}^n (X_i \in H_i)\right) = \prod_{i=1}^n P(X_i \in H_i), \quad H_i \in \mathcal{B}_{d_i}, i = 1, \dots, n.$$

Definizione 0.12 (Insieme quasi certo). Dato uno spazio di probabilità (Ω, \mathcal{F}, P) , si dice che un insieme A è *quasi certo* se esiste $C \in \mathcal{F}$ tale che $C \subseteq A$ e $P(C) = 1$.

Definizione 0.13 (Uguaglianza quasi certa). Due variabili aleatorie X, Y definite sullo spazio di probabilità (Ω, \mathcal{F}, P) sono dette *uguali quasi certamente* se, definito

$$(X = Y) := \{\omega \in \Omega \mid X(\omega) = Y(\omega)\},$$

$(X = Y)$ è un evento quasi certo.

Definizione 0.14 (Proprietà quasi certa). Consideriamo una "proprietà" $\mathcal{P}(\omega)$ la cui validità dipende da $\omega \in \Omega$. Diciamo che \mathcal{P} è *quasi certa* se l'insieme

$$A := \{\omega \in \Omega \mid \mathcal{P}(\omega) \text{ è vera}\}$$

è quasi certo.

Definizione 0.15 (Convergenza quasi certa). Consideriamo uno spazio di probabilità (Ω, \mathcal{F}, P) su cui sono definite una successione di variabili aleatorie $\{X_n\}_{n \geq 0}$ e una variabile aleatoria X a valori in \mathbb{R}^d . Diciamo che $\{X_n\}_{n \geq 0}$ *converge quasi certamente* se la proprietà "esiste $\lim_{n \rightarrow \infty} X_n(\omega)$ " è quasi certa. In particolare diciamo che $\{X_n\}_{n \geq 0}$ converge quasi certamente a X se

$$P(\lim_{n \rightarrow \infty} X_n = X) = 1.$$

Definizione 0.16 (Valore atteso). Sia X una variabile aleatoria integrabile su (Ω, \mathcal{F}, P) . Definiamo il *valore atteso* di X come

$$E[X] := \int_{\Omega} X dP = \int_{\Omega} X(\omega) P(d\omega).$$

Teorema 0.4 (Teorema del calcolo della media). Siano $X : \Omega \rightarrow \mathbb{R}^d$ una variabile aleatoria su (Ω, \mathcal{F}, P) con legge μ_X e $f : \mathbb{R}^d \rightarrow \mathbb{R}^N$ una funzione \mathcal{B}_d -misurabile. Allora $f \circ X \in L^1(\Omega, P)$ se e solo se $f \in L^1(\mathbb{R}^d, \mu_X)$ e in tal caso vale

$$E[f(X)] = \int_{\mathbb{R}^d} f d\mu_X.$$

In particolare, se $\mu_X = \sum_{k=1}^{\infty} p_k \delta_{X_k}$ è una distribuzione discreta allora

$$E[f(X)] = \sum_{k=1}^{\infty} f(x_k) p_k,$$

mentre se μ_X è assolutamente continua con densità γ_x si ha

$$E[f(X)] = \int_{\mathbb{R}^d} f(x) \gamma_X(x) dx.$$

Teorema 0.5 (Legge dei grandi numeri forte). Sia $\{X_n\}_{n \geq 0}$ una successione di variabili aleatorie reali, indipendenti identicamente distribuite in $L^1(\Omega, P)$.

Sia $\mu = E[X_1]$ e definiamo

$$M_n := \frac{1}{n} \sum_{j=1}^n X_j.$$

Si ha che

$$\lim_{n \rightarrow \infty} M_n = \mu \quad P - q.c..$$

Teorema 0.6. Siano $X \in L^1(\Omega, \mathcal{F}, P)$ a valori in \mathbb{R}^d e \mathcal{G} una sotto- σ -algebra di \mathcal{F} . Esiste una variabile aleatoria $Z \in L^1(\Omega, P)$ a valori in \mathbb{R}^d che soddisfa le seguenti proprietà:

1. Z è \mathcal{G} -misurabile;
2. per ogni v.a. W \mathcal{G} -misurabile e limitata, vale $E[ZW] = E[XW]$.

Inoltre due variabili aleatorie Z, Z' con queste proprietà sono uguali quasi certamente.

Definizione 0.17 (Attesa condizionata). Siano $X \in L^1(\Omega, \mathcal{F}, P)$ e \mathcal{G} una sotto- σ -algebra di \mathcal{F} . Se Z soddisfa le proprietà del Teorema 0.6 scriviamo, con un abuso di notazione,

$$Z = E[X \mid \mathcal{G}].$$

Diciamo che Z è una versione dell'attesa condizionata di X a \mathcal{G} . In particolare, se $\mathcal{G} = \sigma(Y)$ con Y v.a. su (Ω, \mathcal{F}, P) , scriviamo

$$Z = E[X \mid Y],$$

invece di $Z = E[X \mid \sigma(Y)]$.

Teorema 0.7 (Proprietà dell'attesa condizionata). *Siano $X, Y \in L^1(\Omega, \mathcal{F}, P)$ due variabili aleatorie reali e siano \mathcal{G}, \mathcal{H} due sotto- σ -algebre di \mathcal{F} . Valgono le seguenti proprietà:*

1. (**Formula della probabilità totale**)

$$E[X] = E[E[X | \mathcal{G}]].$$

2. Se X è \mathcal{G} -misurabile allora

$$X = E[X | \mathcal{G}].$$

3. Se X e \mathcal{G} sono indipendenti allora

$$E[X] = E[X | \mathcal{G}].$$

4. (**Linearità**) $\forall a \in \mathbb{R}$ si ha

$$aE[X | \mathcal{G}] + E[Y | \mathcal{G}] = E[aX + Y | \mathcal{G}].$$

5. (**Monotonia**) Se $P(X \leq Y) = 1$ allora

$$E[X | \mathcal{G}] \leq E[Y | \mathcal{G}],$$

nel senso che se $Z = E[X | \mathcal{G}]$ e $W = E[Y | \mathcal{G}]$ allora $P(Z \leq W) = 1$.

6. Se X è \mathcal{G} -misurabile e limitata, si ha

$$XE[Y | \mathcal{G}] = E[XY | \mathcal{G}].$$

7. (**Proprietà della torre**) Se $\mathcal{H} \subseteq \mathcal{G}$, si ha

$$E[E[X | \mathcal{G}] | \mathcal{H}] = E[X | \mathcal{H}].$$

8. (**Lemma di Freezing**) Siano \mathcal{G}, \mathcal{H} indipendenti, X \mathcal{G} -misurabile e $f = f(x, \omega)$ $(\mathcal{B} \otimes \mathcal{H})$ -misurabile tale che $f(X, \cdot) \in L^1(\Omega, P)$ oppure $f \geq 0$. Allora si ha

$$E[f(X, \cdot) | \mathcal{G}] = F(X) \text{ dove } F(x) := E[f(x, \cdot)],$$

o, con una scrittura più compatta,

$$E[f(X, \cdot) | \mathcal{G}] = E[f(x, \cdot)]|_{x=X}.$$

Proposizione 0.8 (Lemma di Freezing). *Sia \mathcal{G} una sotto- σ -algebra di \mathcal{F} . Se X è \mathcal{G} -misurabile, Y è una v.a. indipendente da \mathcal{G} e f \mathcal{B}_2 -misurabile è tale che $f(X, Y) \in L^1(\Omega, P)$, allora si ha*

$$E[f(X, Y) \mid \mathcal{G}] = F(X) \text{ dove } F(x) := E[f(x, Y)],$$

o, con una scrittura più compatta,

$$E[f(X, Y) \mid \mathcal{G}] = E[f(x, Y)]|_{x=X}.$$

Definizione 0.18 (Processo stocastico a tempo discreto). Sia $I \subseteq \mathbb{N}_0$. Un *processo stocastico a tempo discreto* su (Ω, \mathcal{F}, P) è una famiglia di variabili aleatorie $X = \{X_t\}_{t \in I}$.

Bibliografia

- [B] PIERRE BRÉMAUD. Markov Chains: Gibbs fields, Monte Carlo simulation, and queues. *Springer*, 1999.
- [R] SHELDON M. ROSS. Introduction to probability models, 12ma edizione. *Academic press*, 2019.
- [P] ANDREA PASCUCCI Teoria della probabilità. *Springer-Verlag Italia*, 2020.

Ringraziamenti

Ringrazio il prof. Stefano Pagliarani per avermi guidato nella stesura della tesi, la mia famiglia e tutte le persone che mi hanno supportato durante il percorso di studi.