

Relazione di Laboratorio Computazionale

Alessio Marchetti

Le parti in grassetto sono commenti generici o placeholder perchè QUESTA È UNA BOZZA!

Abstract

In questa relazione prenderemo in considerazione il problema di estrarre campioni di valori casuali data una certa distribuzione di probabilità discreta. Supporremo di sapere generare variabili uniformi sull'intervallo $[0, 1]$, che verranno implementate operativamente come le variabili generate dalla libreria `numpy`.

Una prima soluzione del problema è quella di dividere $[0, 1]$ in intervalli di lunghezza pari a ciascuna componente del vettore di probabilità, e scegliere il risultato in funzione dell'intervallo a cui appartiene una variabile uniforme. Questo presenta diversi inconvenienti: il metodo infatti richiede un numero di somme proporzionale al numero di componenti del vettore di probabilità. Questo potrebbe essere intrattabile quando molto grande. Inoltre spesso il vettore di probabilità è noto solo a meno di un coefficiente di normalizzazione, il cui calcolo richiederebbe di nuovo $O(n)$ somme.

Uno dei metodi più utilizzati per ovviare a questi problemi è il metodo di Monte Carlo (abbreviato spesso con MCMC, Markov Chain Monte Carlo), che consiste nella simulazione di una camminata su una catena di Markov con distribuzione invariante la distribuzione data. Dopo un numero sufficiente di step, la frequenza di visita di un nodo sarà arbitrariamente vicina a quella voluta. In questo caso però il numero di passi necessari ad una determinata distribuzione non è noto a priori ed è di difficile calcolo.

Si andrà dunque a presentare l'algoritmo di Propp-Wilson, una modifica del MCMC che ha il vantaggio di ottenere la distribuzione esatta e di terminare una volta raggiunta questa. Applicheremo tale algoritmo al modello di Ising, una modellizzazione del comportamento magnetico della materia.

1 Il modello di Ising

Sia $G = (V, E)$ un grafo. I vertici andranno a rappresentare i singoli atomi di un materiale, e gli archi indicano quali atomi interagiscono fra loro. Ad ogni atomo viene quindi associato uno spin che può essere $+1$ o -1 . Una configurazione è quindi una funzione $f: V \rightarrow \{+1, -1\}$. A ciascuno di questi modelli si associa

l'energia

$$H(f) = \sum_{(x,y) \in E} f(x)f(y)$$

. Inoltre viene dato un parametro reale del sistema $\beta \geq 0$ detta temperatura inversa. Il modello di ising associa ad ogni configurazione $f \in \{+1, -1\}^V$ la probabilità

$$\pi(f) = \frac{1}{Z} \exp(-\beta H(f))$$

Dove Z è il coefficiente di normalizzazione pari a

$$Z = \sum_{f \in \{+1, -1\}^V} \exp(-\beta H(f))$$

L'obiettivo che ci prefiggiamo è quello di estrarre un campione da $\{+1, -1\}^V$ con probabilità π .

2 Metodo di Monte Carlo

Non so se ha davvero senso questa parte

Come prima cosa, data una distribuzione π su un insieme Ω , cerchiamo una matrice di transizione P per una catena di Markov omogenea che abbia π come misura invariante. Per farlo chiediamo una condizione più forte su P , cioè la reversibilità, ovvero si vuole che $\pi(i)P_{i,j} = \pi(j)P_{j,i}$. Se la matrice P è irriducibile e aperiodica, data una distribuzione iniziale π_0 , la distribuzione di probabilità dopo n passi $\pi_n = \pi_0 P^n$ converge a π

Cerchiamo allora le matrici del tipo $P_{i,j} = A_{i,j}Q_{i,j}$ per $i \neq j$, dove la matrice Q è irriducibile e viene detta matrice generatrice dei candidati, e A è da terminare in modo tale che P abbia le proprietà richieste e $0 \leq A \leq 1$. Le componenti $P_{i,i}$ sono determinate in maniera tale che P sia stocastica. A livello di interpretazione si può pensare che a ogni step si sceglie vertice candidato a cui passare con probabilità dettata da Q e poi si esegue il passaggio con probabilità dettata da A , altrimenti si rimane nel vertice di partenza. Per questo motivo A è la matrice delle probabilità di accettazione.

Generalmente si sceglie A della forma

$$A_{i,j} = \frac{S_{i,j}}{1 + T_{i,j}}$$

con

$$T_{i,j} = \frac{\pi(i)Q_{i,j}}{\pi(j)Q_{j,i}}$$

e S una matrice simmetrica. L'algoritmo di Metropoli-Hastings sceglie

$$A_{i,j} = \min(1, \frac{\pi(j)}{\pi(i)}).$$

In questo modo si determina una P con tutte le proprietà richieste.

3 Il Gibbs sampler

come gestico le traduzioni? Vogliamo specializzare il MCMC ai casi in cui l'insieme degli stati sia un insieme di funzioni (con dominio e codominio finiti). Questa capita nel modello di Ising che studieremo in seguito, per esempio. Abbiamo dunque un processo stocastico a valori in $E = \Lambda^V$ per certi insiemi Λ e V .

Con il Gibbs sampler il processo passa da uno stato $X_n = f$ a $X_{n+1} = g$ con le seguenti regole: sia v un elemento scelto in modo casuale uniforme da V . Allora f e g devono coincidere su $V \setminus \{v\}$. Il punto $g(v)$ assume il valore λ con probabilità $\mathbb{P}[X(v) = \lambda \mid X(V \setminus \{v\}) = f(V \setminus \{v\})]$. **Si verifica che** la distribuzione cercata soddisfa la condizione di reversibilità e che la catena risultante è irriducibile e aperiodica. Dunque si ha la convergenza del metodo di Monte Carlo.

4 Coupling from the past

In questa sezione svilupperemo un metodo per estrarre un campione con una probabilità esatta π , data P una matrice di transizione irriducibile e aperiodica su un insieme finito di stati $S = \{s_1, \dots, s_n\}$ e con probabilità invariante π . Possiamo associare alla catena di Markov definita da P una funzione di transizione

$$f: S \times [0, 1] \longrightarrow S$$

tale che se U è una variabile uniforme sull'intervallo $[0, 1]$, allora

$$\mathbb{P}[f(s_i, U) = s_j] = P_{i,j} \quad \forall s_i, s_j \in S.$$

Questo può essere fatto, per esempio **nel modo ovvio**.

Siano U_n variabili uniformi su $[0, 1]$ per ogni $n \in \mathbb{Z}$. Costruiamo allora delle sequenze $X_m^r(i)$ con $i = 1, \dots, n$, $r \leq 0$ e $m \geq r$ interi nel modo seguente:

$$X_r^r(i) = s_i$$

$$X_{m+1}^r(i) = f(X_m^r(i), U_m).$$

Sia ora

$$\tau^- = \max\{r \mid X_0^r(1) = \dots = X_0^r(n)\}$$

e definiamo $Y = X_0^{\tau^-}(1)$. La condizione con cui è stato definito τ^- è detta coalescenza.

Si verifica che Y ha distribuzione pari a π .

Questo metodo è noto come algoritmo di Propp-Wilson.

Probabilmente questo ci va.

Probabilmente ci vanno anche dei controesempi per dire che il coupling in the future non funziona e che gli U devono essere sempre gli stessi, c'è su Haggstrom pp.81-82.

Mettere i disegnini che sono strabelli?

5 Sandwiching

Il metodo presentato sopra è funzionante e risolve gli scopi che ci eravamo prefissi, tuttavia simulare tutte le n catene di Markov è praticamente impossibile per n grandi. In questa sezione ci occuperemo di migliorare il metodo per renderlo computazionalmente più leggero.

In questo caso supporremo di avere un ordine parziale tra gli stati in S che denoteremo con il simbolo \preceq . Richiediamo che la funzione di transizione rispetti la condizione di monotonia

$$f(s_i, u) \leq f(s_j, u) \quad \forall s_i \preceq s_j$$

e che esistano in S un massimo e un minimo, che senza perdita di generalità chiameremo s_1 e s_n . Si verifica facilmente che la coalescenza di X_0^r si verifica se e solo se $X_0^r(1) = X_0^r(n)$. Inoltre il tempo di coalescenza τ^- è quasi certamente finito.

Forse anche questa cosa si dovrebbe fare...

6 Implementazione del modello di Ising

Andiamo ad utilizzare le tecniche viste al modello di Ising. Innanzi tutto ci ridurremo al caso in cui il grafo è una griglia quadrata di lato l . Sui possibili stati di un grafo, consideriamo la relazione d'ordine definita da: $f \preceq g$ se e solo se per ogni vertice v si ha che $f(v) \leq g(v)$. Con questa relazione si hanno due stati particolari, le due funzioni costanti, che sono il massimo e il minimo. Lo scopo è quello di simulare le catene di Markov che partono da questo stato.

Per farlo prenderemo in considerazione il modello di Gibbs. Ad ogni step di configurazione f , si sceglie in modo casuale uniforme un vertice v , e lo si pone a $+1$ con probabilità

$$\frac{2\beta \sum_{(v,v') \in E} f(v')}{2\beta \sum_{(v,v') \in E} f(v') + 1}.$$

Lo si pone -1 altrimenti. Questa catena di Markov è monotona, infatti i vertici che vengono aggiornati sono gli stessi per la catena che parte dal massimo e quella che parte dal minimo, inoltre vengono aggiornati nello stesso modo. **Per una cosa vista perlando del Gibbs sampler**, la distribuzione stazionaria è quella del modello di Ising.

Viene fornito di seguito il codice per l'implementazione.