Markov Chains and Markov Chain Monte Carlo (MCMC)

Alberto Garfagnini

Università degli studi di Padova

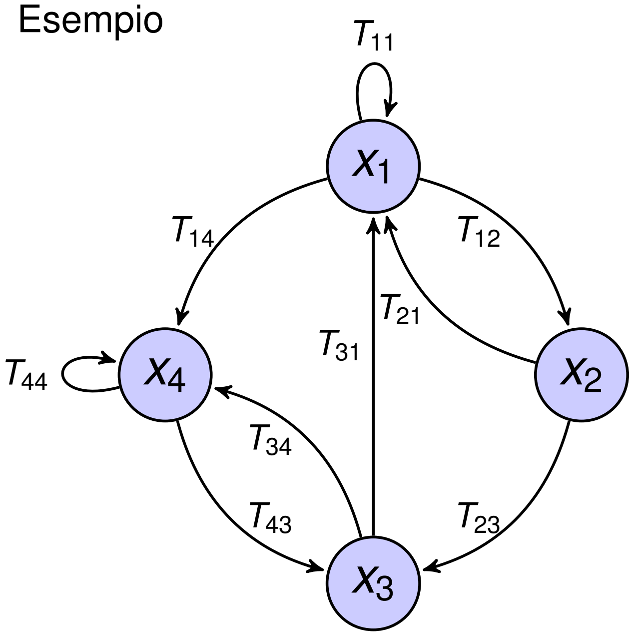
December 11, 2013

Catene di Markov Discrete

- dato un valore x_t del sistema ad un istante di tempo fissato, il sistema evolve ad uno stato x_{t+1} nell'iterazione successiva
- le probabilità di evoluzione del sistema dipendono dallo stato attuale del sistema
- Alcuni esempi di processi probabilistici ricorrenti si trovano in
  - cambiamenti nelle specie delle sequenze del DNA attraverso le mutazioni
  - i passi che determinano il ripiegamento di una proteina
  - l'andamento giornaliero del mercato azionario
- in tutti questi esempi, il sistema deve evolvere tra uno stato (definito ad un istante di tempo) e quello temporalmente successivo in maniera casuale basandosi esclusivamente sullo stato presente e non sull'intera storia dell'evoluzione

Grafo per il Random Walk

- La Catena di Markov può essere rappresentata da un grafo :
  1. i nodi rappresentano gli stati
  2. le frecce le transizioni possibili con le relative probabilità
- Esempio

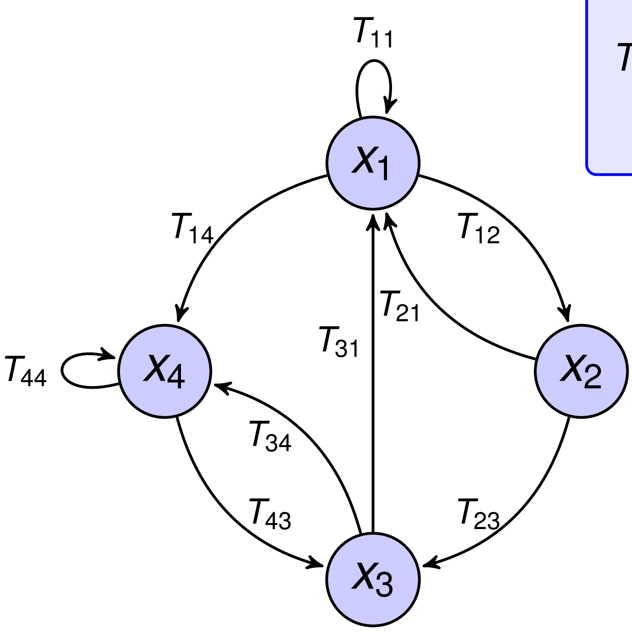


T_{jl} probabilità di transizione $x_j \rightarrow x_l$

Matrice di Transizione

- È possibile associare una matrice di transizione $T = \{T_{ij}\}$

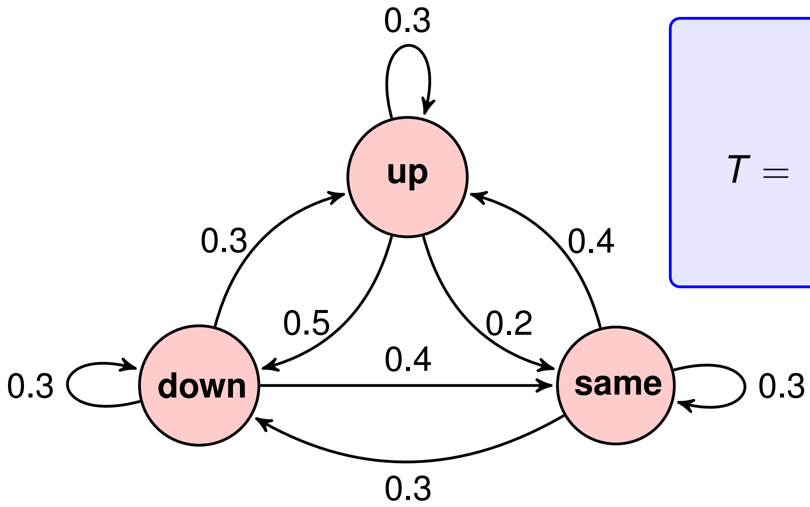
$$T = \begin{pmatrix} T_{11} & T_{12} & 0 & T_{14} \\ T_{21} & 0 & T_{23} & 0 \\ T_{31} & 0 & 0 & T_{34} \\ 0 & 0 & T_{43} & T_{44} \end{pmatrix}$$



La somma delle probabità da uno stato x_i agli altri stati deve essere normalizzata $\sum_{j=1}^{N} T_{ij} = 1$

Es: modello stocastico di mercato azionario

- In relazione all'indice della borsa odierno, nella giornata di domani, il mercato potrà avere un andamento al rialzo (up), ribasso (down), oppure rimanere stazionario (same)



$$T = \begin{array}{c} \begin{array}{ccc} \text{up} & \text{same} & \text{down} \end{array} \\ \begin{pmatrix} 0.3 & 0.2 & 0.5 \\ 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{pmatrix} \end{array}$$

$$\sum_{j=1}^{N} T_{ij} = 1$$

Evoluzione del sistema

- Definiamo con

$$\boldsymbol{\pi}^{(k)} = [\pi_1^{(k)} \pi_2^{(k)} \dots \pi_n^{(k)}]$$

- un vettore riga, dove $\pi_j^{(k)}$ indica la probabilità che la variabile casuale x si trovi nello stato x_j allo step k
- la probabilità che la catena evolva nello stato x_i allo step $k+1$ è

$$\pi_i^{(k+1)} = \sum_{j=1}^{n} \pi_j^{(k)} T_{ji} \tag{1}$$

- che si può scrivere in forma matriciale come

$$\boldsymbol{\pi}^{(k+1)} = \boldsymbol{\pi}^{(k)} \boldsymbol{T} \tag{2}$$

Evoluzione del sistema (2)

- Dall'eq (2) si può ricavare la relazione

$$\boldsymbol{\pi}^{(k)} = \boldsymbol{\pi}^{(k-1)}\boldsymbol{T} = (\boldsymbol{\pi}^{(k-2)}\boldsymbol{T})\boldsymbol{T} = \boldsymbol{\pi}^{(k-2)}\boldsymbol{T}^2 \tag{3}$$

- ed inferire che

$$\boldsymbol{\pi}^{(k)} = \boldsymbol{\pi}^{(0)}\boldsymbol{T}^k \tag{4}$$

- con il crescere di k $\boldsymbol{\pi}^{(k)}$ tende ad una distribuzione stazionaria, o alla PDF sulla quale la matrice di transizione è basata:

$$f_i = \lim_{n\to\infty} \pi_i^{(n)}$$

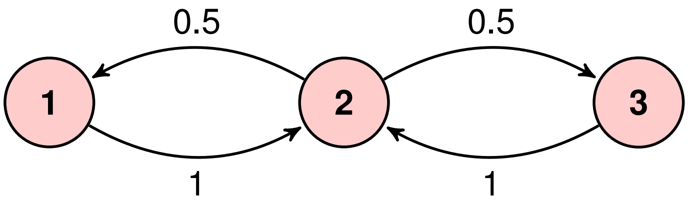
Distribuzione invariante

- Per una distribuzione invariante avremo

$$\boldsymbol{\pi} T = \boldsymbol{\pi}$$

- le condizioni che deve soddisfare la matrice di transizione $\boldsymbol{T}$ affinchè il sistema tenda ad una distribuzione invariante sono:

1) $\boldsymbol{T}$ è irriducibile : nella rappresentazione del grafo, ogni vertice è collegato direttamente ad ogni altro vertice
2) $\boldsymbol{T}$ è aperiodica : non è possibile trovare un intero $d > 1$ tale che il numero di iterazioni necessarie per ritornare ad uno stato fissato sia multiplo di d

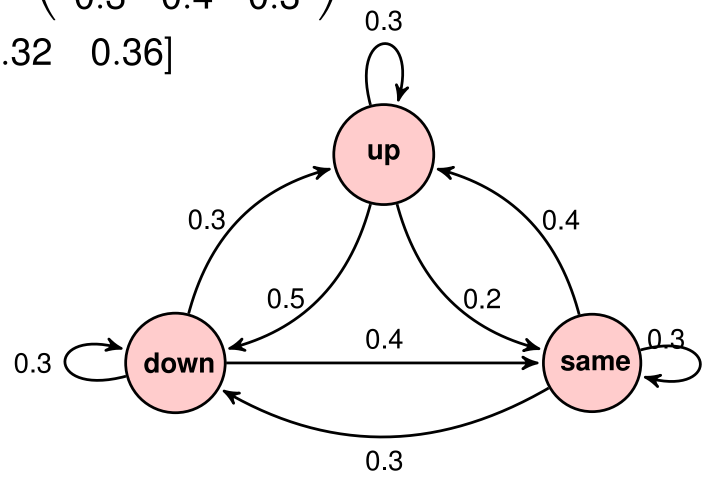


Catena periodica con periodo $d = 2$

Modello stocastico di mercato azionario (1)

- Supponendo che il mercato azionario sia in rialzo (UP), calcoliamo le probabilità di evoluzione al giorno successivo

$$\begin{aligned}
\pi^{(1)} &= \pi^{(0)} T \\
&= [1 \quad 0 \quad 0] \begin{pmatrix} 0.3 & 0.2 & 0.5 \\ 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \end{pmatrix} \\
&= [0.32 \quad 0.32 \quad 0.36]
\end{aligned}$$



Modello stocastico di mercato azionario (2)

- Continuando ad evolvere il sistema ai giorni successivi si ottiene:

$$\begin{aligned}
\pi^{(2)} &= \pi^{(0)} T^2 \\
&= [0.332 \quad 0.304 \quad 0.364] \\
\pi^{(3)} &= [0.3304 \quad 0.303 \quad 0.3664] \\
\pi^{(4)} &= [0.33032 \quad 0.3036 \quad 0.36608] \\
&\dots \\
\pi^{(9)} &= [0.330357 \quad 0.303571 \quad 0.366072]
\end{aligned}$$

- il sistema si stabilizza con poche iterazioni e tende al limite della distribuzione invariante
- risolvendo l'equazione $\boldsymbol{\pi T} = \boldsymbol{\pi}$ si ottiene:

$$\pi = \left[\frac{37}{112} \quad \frac{34}{112} \quad \frac{41}{112}\right] \sim [0.33035714 \quad 0.30357143 \quad 0.36607143]$$

Markov Chain Monte Carlo

- Idea suggerita da N. Metropolis nel 1953 nell'articolo "Equations of state calculations by fast computing machines" : tratta il moto casuale di atomi e molecole ed il calcolo delle loro energie cinetiche e perfezionata da Hastings nel 1970
- L'algoritmo proposto da Metropolis è nella top ten degli algoritmi più citati con la maggiore influenza nello sviluppo della scienza e ingegneria
- idea di base:
  - costruire una Catena di Markov la cui distribuzione invariante è la distribuzione desiderata dalla quale campionare
  - Se f è una densità di probabilità discreta, con valori nel dominio Ω, i campionamenti da f saranno approssimati da una Catena di Markov in Ω
  - La generazione dello stato successivo è un processo a due fasi: un valore proposto (estratto da una distribuzione), seguito da accettanza o reiezione

Markov Chain Monte Carlo: algoritmo

- $f(x)$, definita in $D = [a, b]$, è la PDF in esame

1. Scegliere una qualsiai PDF $h(u|x)$ definita in D
2. Specificare un valore iniziale D, x_i
3. Campionare un numero u_i da h utilizzando una procedura appropriata (per esempio se la CDF di h è nota analiticamente, usare il metodo dell'inversione)
4. Calcolare la grandezza R, chiamato rapporto di Metropolis o rapporto di Hastings e definita come

$$R(x_i, u_i) = \frac{f(u_i)}{f(x_i)} \frac{h(x_i|u_i)}{h(u_i|x_i)}$$

5. Porre $\alpha(x_i, u_i) = \min[1, R(x_i, u_i)]$
6. generare un numero $\rho \sim U(0, 1)$ ed accettare u_i con probabilità α o rigettarlo con probabilità $1 - \alpha$ poichè

$$x_{i+1} = \begin{cases} u_i & \text{if} \rho \leq \alpha \\ x_i & \text{if} \rho > \alpha \end{cases}$$

Esempio: calcolo di un integrale con MCMC

- Calcolare l'integrale

$$I = \int_0^\infty (x - \alpha\beta)^2 f(x) dx$$

- con

$$f(x) = \frac{1}{\Gamma(\alpha)} \left(\frac{x}{\alpha}\right)^{\alpha-1} \exp(-x/\beta)$$

- e $\alpha = 2$, $\beta = 1$.

Esempio: calcolo di un integrale con MCMC

- Si scelga come PDF di prova $h(u) = \frac{1}{2} \exp(-x/2)$
- e si parta da $x_i = 1$ per $i = 1$.
- campionando da h avremo $u_i = -2 \log \rho$ con $\rho \in U(0, 1)$.
- il rapporto di Metropolis è

$$R = \frac{f(u_i)h(x_i)}{f(x_i)u(x_i)} = \left(\frac{u_i}{x_i}\right)^{\alpha-1} e^{(u_i - x_i)/\beta} e^{(x_i - u_i)/2}$$

- la dinamica della catena di Markov è regolata da

$$x_{i+1} = \begin{cases} u_i & \text{if} \quad \rho \leq \min(1, R) \\ x_i & \text{if} \quad \rho > \min(1, R) \end{cases}$$

- Una stima dell'integrale è $I_N = \frac{1}{N} \sum_{i=1}^N (x_i - \alpha\beta)^2 = \frac{1}{N} \sum_{i=1}^N (x_i - 2)^2$