

Il fenomeno del cutoff nelle catene di Markov

Candidato: Diego Stucchi

Relatore: prof. Francesco Caravenna

Università degli studi di Milano-Bicocca

29-11-2016

1 Catene di Markov

2 Il fenomeno del cutoff

3 Stime dall'alto per il t_{mix}

1 Catene di Markov

2 Il fenomeno del cutoff

3 Stime dall'alto per il t_{mix}

Catena di Markov

Sia Ω un insieme finito (detto *spazio degli stati*)

Catena di Markov

Sia Ω un insieme finito (detto *spazio degli stati*)

Sia $(X_n)_{n \in \mathbb{N}}$ una successione di variabili aleatorie a valori in Ω

Catena di Markov

Sia Ω un insieme finito (detto *spazio degli stati*)

Sia $(X_n)_{n \in \mathbb{N}}$ una successione di variabili aleatorie a valori in Ω
Tale successione si dice *catena di Markov* (omogenea) se

$$P(X_{n+1} = j | X_n = i) = P(i, j)$$

dove $P \in \mathbb{R}^{|\Omega| \times |\Omega|}$ è la *matrice di transizione*

Catena di Markov

Sia Ω un insieme finito (detto *spazio degli stati*)

Sia $(X_n)_{n \in \mathbb{N}}$ una successione di variabili aleatorie a valori in Ω
Tale successione si dice *catena di Markov* (omogenea) se

$$P(X_{n+1} = j | X_n = i) = P(i, j)$$

dove $P \in \mathbb{R}^{|\Omega| \times |\Omega|}$ è la *matrice di transizione*

Una probabilità si dice *invariante* se

$$\sum_{i \in \Omega} \pi_i P(i, j) = \pi_j$$

Catena di Markov

Catena *irriducibile* = partendo da uno stato si può raggiungere qualsiasi altro stato

Catena di Markov

Catena *irriducibile* = partendo da uno stato si può raggiungere qualsiasi altro stato

Teorema (Convergenza all'equilibrio)

(X_t) catena irriducibile aperiodica su Ω finito
 π distribuzione invariante

$$\lim_{n \rightarrow +\infty} P^n(i,j) = \pi_j \quad \forall i, j \in \Omega$$

Due esempi di catene di Markov

Introduciamo due modelli di catene di Markov associate al rimescolamento di un mazzo di carte.

Due esempi di catene di Markov

Introduciamo due modelli di catene di Markov associate al rimescolamento di un mazzo di carte.

Sia X_t l'ordine delle carte dopo t mescolamenti

$$X_t \in S_n$$

Due esempi di catene di Markov

Introduciamo due modelli di catene di Markov associate al rimescolamento di un mazzo di carte.

Sia X_t l'ordine delle carte dopo t mescolamenti

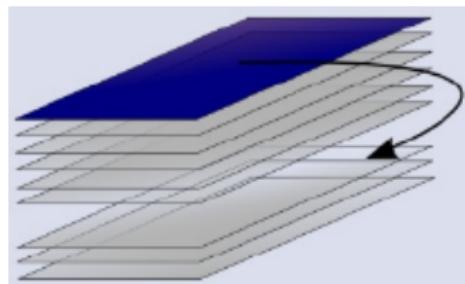
$$X_t \in S_n$$

Descrivo X_t come un'opportuna catena di Markov in cui il metodo di mescolamento caratterizza P

Top-to-random

Il *top-to-random* consiste nel prendere la carta in cima al mazzo e inserirla in una posizione casuale

Questo metodo è molto lento



Mescolamento all'americana



Il mescolamento all'americana consiste nel tagliare il mazzo e nel riunire le due parti in modo casuale mantenendo l'ordine relativo delle carte.

Questo metodo è nettamente più efficace del precedente

Quesito fondamentale

Obiettivo

Quante volte devo mischiare il mazzo di carte per ottenere un mazzo ben mescolato?

1 Catene di Markov

2 Il fenomeno del cutoff

3 Stime dall'alto per il t_{mix}

Una catena irriducibile e aperiodica su Ω finito converge.

Convergenza

Quanto in fretta converge alla probabilità invariante?

Definizione (Distanza in variazione totale)

μ e ν probabilità su Ω finito

$$\|\mu - \nu\|_{TV} = \max_{A \subset \Omega} |\mu(A) - \nu(A)|$$

Una misura della convergenza

Definizione (Distanza dalla stazionarietà)

(X_n) catena su Ω finito, P matrice di transizione
 π probabilità invariante, $P^t(x, \cdot)$ distribuzione di X_t se $X_0 = x$

$$d(t) := \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV}$$

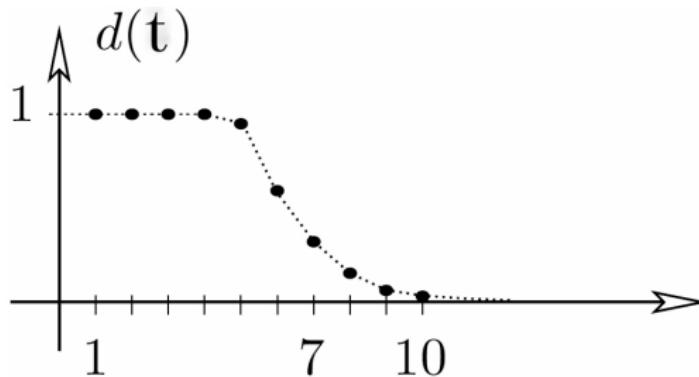
$d(t)$ assume valori tra 0 e 1

Riformulo il quesito

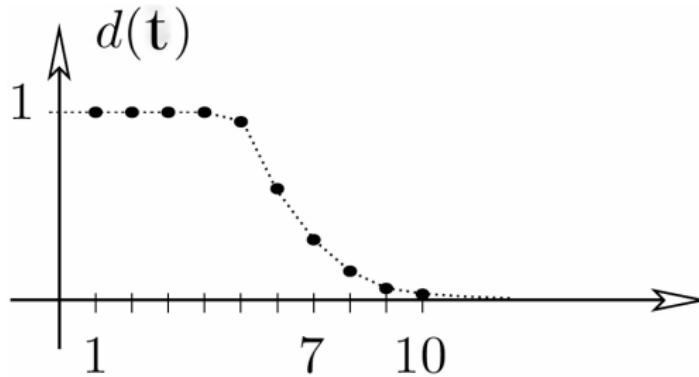
Obiettivo

Quanto deve essere grande t affinché la distribuzione di X_t sia abbastanza vicina alla distribuzione invariante?

Alcune catene di Markov convergono in modo particolare
Il passaggio da $d(t) \sim 1$ a $d(t) \sim 0$ è brusco



Alcune catene di Markov convergono in modo particolare
Il passaggio da $d(t) \sim 1$ a $d(t) \sim 0$ è brusco



Definizione (Tempo di mescolamento)

$$t_{mix} := \min t : d(t) \leq \frac{1}{3}\}$$

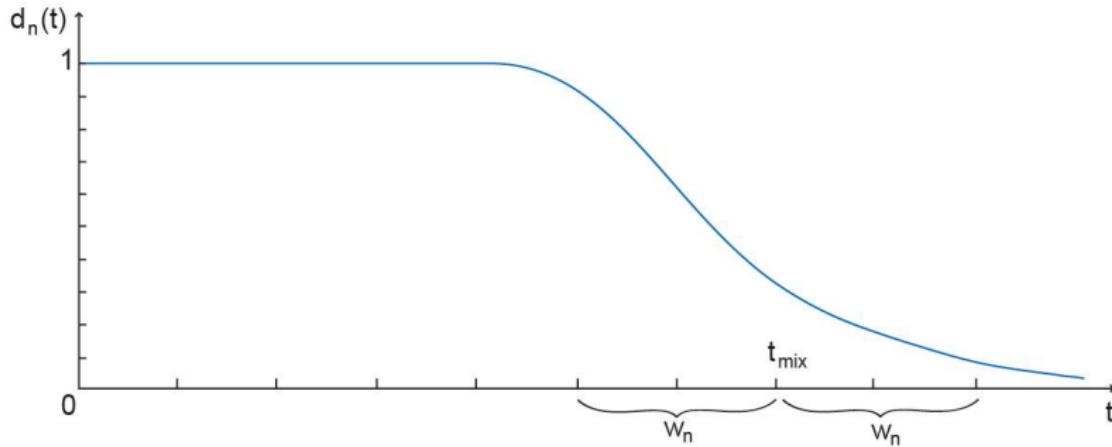
Il cutoff

$(X_t^{(n)})$ famiglia di catene di Markov parametrizzate da n
 $d_n(t)$ distanza per l' n -esima catena, $t_{mix}^{(n)}$ tempo di mescolamento

Definizione

$(X_t^{(n)})$ ha cutoff se

$$\lim_{n \rightarrow \infty} d_n(ct_{mix}^{(n)}) = \begin{cases} 1 & \text{se } c < 1 \\ 0 & \text{se } c > 1 \end{cases}$$

Il grafico di $d_n(t)$ 

Finestra di cutoff

Un cutoff ha una finestra di dimensione $\{\omega_n\}$ se $\omega_n = o(t_{mix}^{(n)})$ e

$$\lim_{\alpha \rightarrow +\infty} \liminf_{n \rightarrow \infty} d_n(t_{mix}^{(n)} - \alpha \omega_n) = 1$$

e

$$\lim_{\alpha \rightarrow +\infty} \limsup_{n \rightarrow \infty} d_n(t_{mix}^{(n)} + \alpha \omega_n) = 0$$

Come si nota dal grafico il crollo brusco avviene attorno a t_{mix} con una finestra ω_n

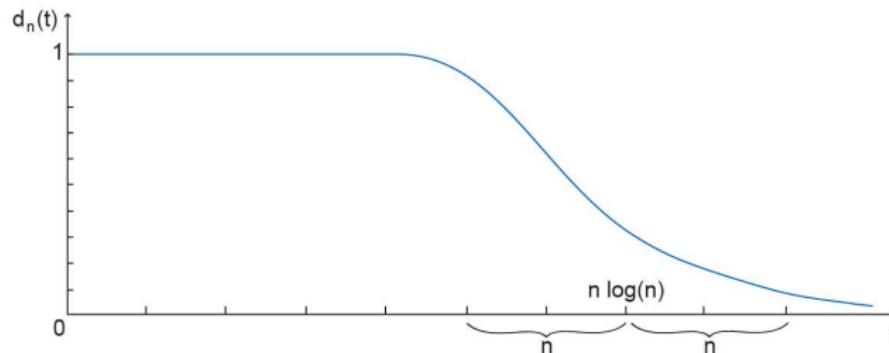
Top-to-random

Teorema

Il top-to-random ha un cutoff con tempo di mescolamento

$$t_{mix}^{(n)} = n \log n$$

con finestra $\omega_n = n$



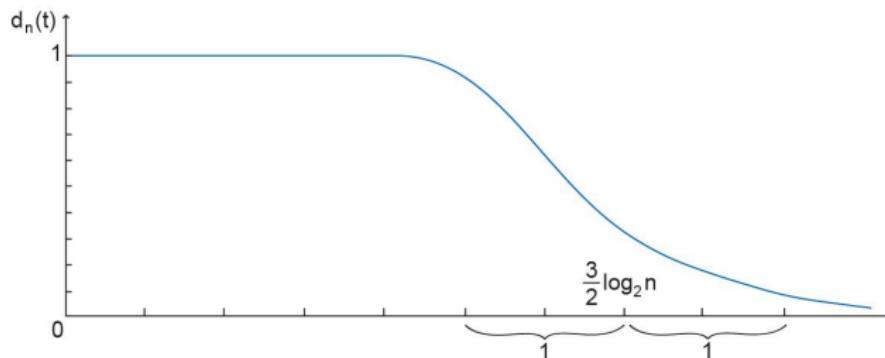
Riffle shuffle

Teorema

Il riffle shuffle ha un cutoff con tempo di mescolamento

$$t_{mix}^{(n)} = \frac{3}{2} \log_2 n$$

con finestra $\omega_n = 1$



1 Catene di Markov

2 Il fenomeno del cutoff

3 Stime dall'alto per il t_{mix}

Tempi aleatori

Definizione (Tempo di arresto)

$$\{\tau = n\} \in \sigma(X_1, \dots, X_n)$$

Tempi aleatori

Definizione (Tempo di arresto)

$$\{\tau = n\} \in \sigma(X_1, \dots, X_n)$$

Nel top-to-random definisco τ_{top} = istante in cui la carta originariamente sul fondo viene reinserita nel mazzo

τ_{top} è un tempo di arresto

Tempi aleatori

Definizione (Tempo stazionario forte)

Tempo di arresto τ tale che:

X_τ ha distribuzione invariante

X_τ è indipendente da τ

Tempi aleatori

Definizione (Tempo stazionario forte)

Tempo di arresto τ tale che:

X_τ ha distribuzione invariante

X_τ è indipendente da τ

È facile mostrare che nel top-to-random $X_{\tau_{top}}$ ha distribuzione uniforme

τ_{top} è un tempo stazionario forte

Stima di t_{mix}

Teorema

τ tempo stazionario forte

$$d(t) \leq \max_{x \in \Omega} P_x(\tau > t)$$

Distribuzione di τ_{top}

$$\tau_{top} = \tau_1 + \tau_2 + \cdots + \tau_{n-1} + 1$$

Distribuzione di τ_{top}

$$\tau_{top} = \tau_1 + \tau_2 + \cdots + \tau_{n-1} + 1$$

τ_i = numero di rimescolamenti necessari alla carta originariamente sul fondo per salire dall' i -esima posizione dal fondo alla $(i+1)$ -esima

Distribuzione di τ_{top}

$$\tau_{top} = \tau_1 + \tau_2 + \cdots + \tau_{n-1} + 1$$

τ_i = numero di rimescolamenti necessari alla carta originariamente sul fondo per salire dall' i -esima posizione dal fondo alla $(i+1)$ -esima

$$\tau_i \sim Geo\left(\frac{1}{n}\right)$$

Distribuzione geometrica

$$p \in (0, 1)$$

Distribuzione geometrica

$$p \in (0, 1)$$

$$Z \sim Geo(p)$$

$$P(Z = k) = (1 - p)^{k-1} p$$

Distribuzione geometrica

$$\textcolor{red}{X} \quad 1 - p \quad (1)$$

$$p \in (0, 1)$$

$$Z \sim Geo(p)$$

$$P(Z = k) = (1 - p)^{k-1} p$$

Distribuzione geometrica

$$p \in (0, 1)$$

$$\begin{array}{ll} \text{X} & 1 - p \quad (1) \\ \text{X} & 1 - p \quad (2) \end{array}$$

$$Z \sim Geo(p)$$

$$P(Z = k) = (1 - p)^{k-1} p$$

Distribuzione geometrica

$$p \in (0, 1)$$

$$\begin{array}{ll} \text{X} & 1 - p \quad (1) \\ \text{X} & 1 - p \quad (2) \end{array}$$

:

$$Z \sim Geo(p)$$

$$P(Z = k) = (1 - p)^{k-1} p \quad \begin{array}{ll} \text{X} & 1 - p \quad (k-1) \end{array}$$

Distribuzione geometrica

$$p \in (0, 1)$$

X $1 - p$ (1)
X $1 - p$ (2)

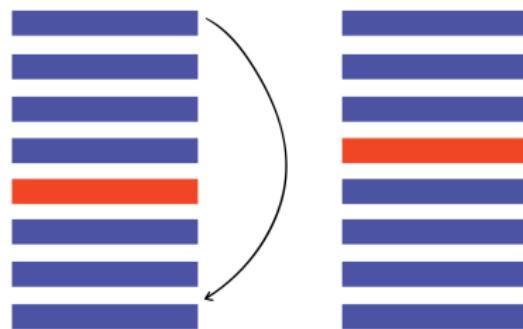
⋮

$$Z \sim Geo(p)$$

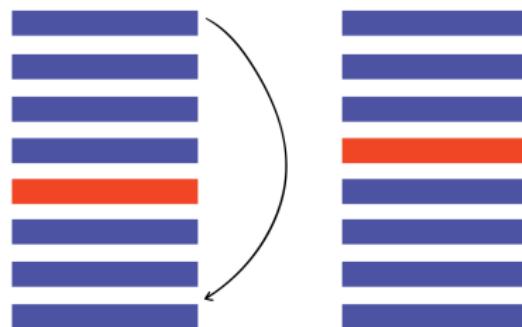
$$P(Z = k) = (1 - p)^{k-1} p$$

X $1 - p$ (k-1)
✓ p (k)

Analisi del top-to-random



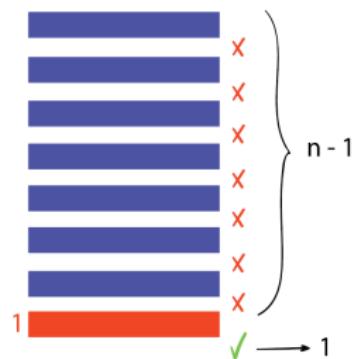
Analisi del top-to-random

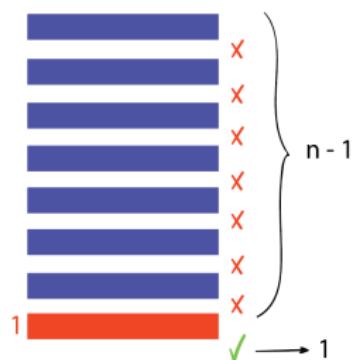


Carta sale di una posizione

SE

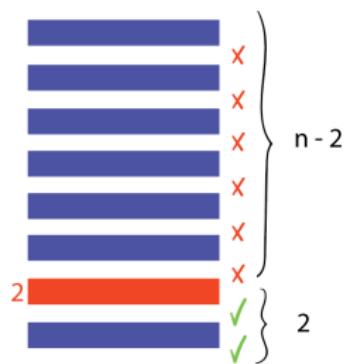
Top card finisce sotto

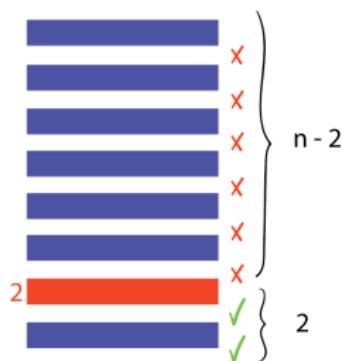
Distribuzione τ_1 

Distribuzione τ_1 

$$P(\text{carta rossa sale}) = \frac{1}{n}$$

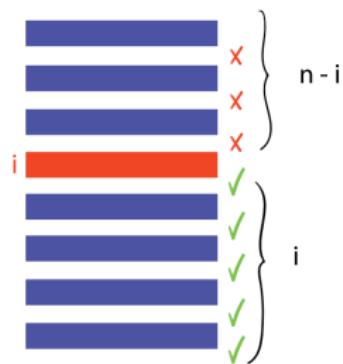
$$\tau_1 \sim Geo\left(\frac{1}{n}\right)$$

Distribuzione τ_2 

Distribuzione τ_2 

$$P(\text{carta rossa sale}) = \frac{2}{n}$$

$$\tau_2 \sim Geo\left(\frac{2}{n}\right)$$

Distribuzione τ_i 

$$P(\text{carta rossa sale}) = \frac{i}{n}$$

$$\tau_i \sim Geo\left(\frac{i}{n}\right)$$

τ_{top} somma di variabili aleatorie geometriche indipendenti

τ_{top} somma di variabili aleatorie geometriche indipendenti



$$P(\tau_{top} > n \log n + cn) \leq e^{-c}$$

τ_{top} somma di variabili aleatorie geometriche indipendenti



$$P(\tau_{top} > n \log n + cn) \leq e^{-c}$$



$$d(n \log n + cn) \leq e^{-c}$$

Stime per il riffle shuffle

Teorema (Bayer-Diaconis)

Riffle shuffle di n carte, P matrice di transizione
 U distribuzione uniforme, $t = \frac{3}{2} \log_2 n + \theta$

$$d(t) \rightarrow 1 - 2\Phi\left(\frac{-2^{-\theta}}{4\sqrt{3}}\right) \quad \text{per } n \rightarrow \infty$$

con

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

Grazie per l'attenzione

Modello Riffle Shuffle

Mazzo diviso in M e $n - M$ carte

$$M \sim \text{Bin}(n, \frac{1}{2})$$

La distribuzione che si ottiene è

$$Q(\sigma) = \begin{cases} \frac{n+1}{2^n} & \text{se } \sigma = id \\ \frac{1}{2^n} & \text{se } \sigma \text{ ha esattamente due sequenze ascendenti} \\ 0 & \text{altrimenti} \end{cases}$$