Legge dei Grandi Numeri

Ricordiamo che se $(X_n)_{n\geq 1}$ sono v. a. i.i.d. reali con media finita μ allora vale la Legge dei Grandi Numeri:

$$\lim_{n\to+\infty} \mathbb{P}(|\overline{X_n}-\mu|\geq \varepsilon)=0 \quad \forall \varepsilon>0$$

con $\overline{X_n} := \frac{1}{n}\sum_{i=1}^{n} X_i$

La dimostrazione é basata, se $Var(X_i)=\sigma^2<+\infty$, sul fatto che $Var(\overline{X_n})=\frac{\sigma^2}{n} \xrightarrow[n\to+\infty]{} 0$

Teorema (Legge dei Grandi Numeri per le CM):
Sia $(X_n)_{n\geq 0}$ una CM irriducibile su E finito, e si denoti con π la sua distribuzione stazionaria. Sia $f: E \to \mathbb{R}$ e poniamo $\pi[f] := \sum_{x\in E} f(x)\pi(x)$. Allora vale la seguente:

$$\boxed{\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i)-\pi[f]\right|\geq \varepsilon\right] \xrightarrow[n\to+\infty]{} 0}$$

Dim.
Si basa sul calcolo di $Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$. Si ha:

$$Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)=\frac{1}{n^2}\left(\sum_{i=1}^{n} Var(f(X_i)) + 2\sum_{i,j=1}^{n} Cov(f(X_i), f(X_j))\right)$$

Notiamo che $\sum_{i=1}^{n} Var(f(X_i)) \sim_{+\infty} \mathcal{O}(n)$. Si dimostra, se la catena é a-periodica usando il Teorema di Convergenza all'Equilibrio, che $\exists\, C,c>0$ t.c. $Cov(f(X_i), f(X_j)) \leq Ce^{-c(j-i)}$

$\Rightarrow \sum_{i,j=1}^{n} Cov(f(X_i), f(X_j)) \sim_{+\infty} \mathcal{O}(n)$

Pertanto, $Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \xrightarrow[n\to+\infty]{} 0$

$\Rightarrow$ per la disuguaglianza di Chebyshev si ha:

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \approx \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(f(X_i))$$

Per il Teorema di convergenza si ha:

$$\mathbb{E}(f(X_n)) \xrightarrow[n\to+\infty]{} \pi[f]$$

Teorema di Convergenza

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n} f(X_i) \approx \pi[f]$$

Nel caso di CM periodiche di periodo d ci si riduce al caso a-periodico usando il fatto che le CM $Y_n^{(i)} := X_{nd+i}$, $i = 0, \dots, d-1$ sono irriducibili e a-periodiche

$\square$

Markov Chain Monte Carlo

Sia E un insieme finito, $f: E \longrightarrow \mathbb{R}$, π una distribuzione su E. Vogliamo calcolare $\pi[f] := \sum_{x \in E} f(x) \pi(x)$.
Se $|E| \gg 1$ può essere conveniente (o addirittura necessario) approssimare $\pi[f] \simeq \frac{1}{n} \sum_{i=1}^{n} f(X_i)$ dove:

1) X_i sono i.i.d. con distribuzione π (NON EFFICIENTE !!!)
2) $(X_n)_{n \geq 0}$ forma una CM irriducibile con distribuzione π

Applicazioni

1) Statistica Bayesiana:

Consideriamo un modello statistico, cioè una famiglia $f(x, \theta)$ di densità su $\mathbb{R}$ dipendenti da un parametro $\theta \in S$. Assumiamo $|S| < +\infty$. In statistica Bayesiana si assume una distribuzione "a priori" per il parametro θ, $\pi(\theta)$. Se è disponibile un insieme di dati $\vec{x} = (x_1, \dots, x_n)$ la distribuzione "a posteriori" di θ è data da

$$\pi_{\vec{x}}(\theta) = \frac{\pi(\theta) \prod_{i=1}^{n} f(x_i, \theta)}{\sum_{\gamma \in S} \pi(\gamma) \prod_{i=1}^{n} f(x_i, \gamma)}$$

Uno "stimatore Bayesiano" per $g(\theta)$ (con $g: S \longrightarrow \mathbb{R}$) è dato da $\sum_{\theta \in S} g(\theta) \pi_{\vec{x}}(\theta)$. Per calcolare tale media si cerca una CM $(X_n)_{n \geq 0}$ con distr. staz. $\pi_{\vec{x}}$ e si approssima la media precedente con $\frac{1}{n} \sum_{i=1}^{n} g(X_n)$

2) Problemi di ottimizzazione "duri":

Sia E finito, $F: E \longrightarrow \mathbb{R}$. Vogliamo trovare un minimo di F, ovvero $x^* \in E$ t.c. $F(x^*) \leq F(x)$ $\forall x \in E$. Sia $\beta \gg 1$ e definiamo $\pi_\beta(x) := \frac{e^{-\beta F(x)}}{Z_\beta}$ con $Z_\beta := \sum_{x \in E} e^{-\beta F(x)}$
$\Rightarrow$ se β è grande, π_β è concentrata vicino ai minimi di F, ovvero:

$$\pi_\beta(\{x : F(x) \leq \min F + \varepsilon\}) \xrightarrow[\beta \to +\infty]{} 1$$

Quindi, se $X \sim \pi_\beta$, con alta probabilità X é "quasi" un minimo. Se (X_n) é una CM irriducibile e a-periodica con distr. staz. π_β allora, per n grande, X_n é con alta probabilità "quasi" un minimo

OBBIETTIVO:

assegnata una distr. π su E finito, trovare una matrice Q t.c. $\pi Q = \pi$. Studiamo 2 metodi:

1) ALGORITMO DI METROPOLIS:

Si assegna una matrice stocastica ψ "di riferimento" t.c. $\psi(x,y) = \psi(y,x)$ (simmetrica). ψ seleziona le transizioni possibili (quelle per cui $\psi(x,y) > 0$). Assumiamo però che ψ sia irriducibile. Notare che la distr. reversibile per ψ é la distr. uniforme su E. Vogliamo trovare una perturbazione Q di ψ t.c. $\pi(x) Q(x,y) = \pi(y) Q(x,y)$. Poniamo come incognite $A_{x,y} \in (0,1]$ e poniamo $Q(x,y) = A_{x,y} \cdot \psi(x,y)$ $\forall x \neq y$ (fuori dalla diagonale). Definiamo $Q(x,x) = 1 - \sum_{y \neq x} A_{x,y} \cdot \psi(x,y) \in [0,1]$ dato che $\sum_{y \neq x} A_{x,y} \cdot \psi(x,y) \leq \sum_{y \neq x} \psi(x,y) \leq 1$.

OSSERVAZIONE

$\forall x \neq y$ $Q(x,y) > 0 \Leftrightarrow \psi(x,y) > 0$, quindi Q é irriducibile.

π deve essere reversibile:

$$\forall x,y \quad \pi(x) A_{x,y} \psi(x,y) = \pi(y) A_{y,x} \psi(y,x)$$

$$\Rightarrow \pi(x) A_{x,y} = \pi(y) A_{y,x}$$

ψ simmetrica

Ci sono molte soluzioni di tale sistema di equazioni. Scegliamo $A_{x,y} = \min\left\{1, \frac{\pi(y)}{\pi(x)}\right\} = 1 \wedge \frac{\pi(y)}{\pi(x)}$. Tale A soddisfa l'equazione precedente:

$$\pi(x) A_{x,y} = \pi(x) \wedge \pi(y) = \pi(y) A_{y,x}$$

ESEMPIO:

Sia $G = (V,E)$ un grafo connesso, non diretto e regolare (ogni vertice ha grado d). Se $(x,y) \in E$, scriviamo $x \leftrightarrow y$

Sia $F : V \to \mathbb{R}$. Vogliamo trovare $\min_V F$. Definiamo:

$$\pi_\beta(x) = e^{-\beta F(x)} \cdot \frac{1}{Z_\beta} \quad \text{con} \quad Z_\beta = \sum_{y \in V} e^{-\beta F(y)}$$

Scegliamo come T la matrice di transizione della passeggiata aleatoria su G:

$$T(x,y) = \begin{cases} \frac{1}{d} & x \leftrightarrow y \wedge x \neq y \\ 0 & \text{altrimenti} \end{cases}$$

$$\Rightarrow A_{x,y} = 1 \wedge \frac{\pi(y)}{\pi(x)} = 1 \wedge e^{-\beta F(y) + \beta F(x)} = 1 \wedge e^{\beta(F(x) - F(y))}$$
$$= \exp\left[\beta(F(x) - F(y))^+\right]$$

Quindi:

$$Q(x,y) = \begin{cases} \frac{1}{d} \exp\left[\beta(F(x) - F(y))^+\right] & \text{se } x \leftrightarrow y \\ 1 - \sum_{y \neq x} Q(x,y) & \text{se } x = y \\ 0 & \text{altrimenti} \end{cases}$$

Come abbiamo visto in precedenza, questa dinamica é usata per minimizzare F scegliendo $\beta \ggg 1$. É possibile usare tale algoritmo facendo variare β nel tempo, ottenendo così una CM NON omogenea. Si dimostra che se $\beta_n \xrightarrow[n \to +\infty]{} +\infty \sim \log n$ allora la CM X_n converge quasi certamente ad un minimo di F. Questa variante non omogenea é detta simulated annealing.

2) GIBBS SAMPLER

Tale algoritmo si applica se lo spazio degli stati E é sottoinsieme dello spazio prodotto C^V con C, V finiti.

Sia $x \in E$, $i \in V$, definiamo:

$$\Omega(x,i) = \{ y \in E \text{ t.c. } y_s = x_s \ \forall s \neq i \}$$

La dinamica é la seguente:

1) se x é lo stato corrente, si sceglie casualmente $i \in V$
2) si sceglie $y \in \Omega(x,i)$ con probabilità $\frac{\pi(y)}{\sum_{z \in \Omega(x,i)} \pi(z)}$.

Ciò corrisponde alla matrice di transizione:

$$Q(x,y) = \begin{cases} \frac{1}{|V|} \frac{\pi(y)}{\sum_{z \in \Omega(x,i)} \pi(z)} & \text{se } y \in \Omega(x,i), y \neq x \\ 0 & \text{se } \nexists i \text{ t.c. } y \in \Omega(x,i) \\ 1 - \sum_{y \neq x} Q(x,y) & \text{se } x = y \end{cases}$$

PROPOSIZIONE:

π é reversibile per Q

DIM.:

$x \neq y$:

Se $y \notin \Omega(x,i)$ per qualche i allora $Q(x,y) = 0 \Rightarrow$
$x \notin \Omega(y,i)$ per qualche $i \Rightarrow Q(x,y) = 0$ quindi:

$$\pi(x)\,Q(x,y) = \pi(y)\,Q(y,x)$$

Se $y \in \Omega(x,i)$ $(\Leftrightarrow x \in \Omega(x,i))$ si ha:

$$\pi(x)\,Q(x,y) = \pi(x)\,\frac{1}{|V|}\,\frac{\pi(y)}{\sum_{z \in \Omega(x,i)} \pi(z)}$$

$\square$

ESEMPIO (Modello di Ising):

Sia $G = (V, E)$ grafo, $S = \{-1, 1\}^V$ sp. degli stati.

$$\pi_\beta(x) = \frac{e^{-\beta H(x)}}{Z_\beta} \quad \text{con} \quad H(x) = -\sum_{i,s = i \leftrightarrow s} x_i x_s, \quad \beta > 0$$

Gibbs sampler:

$$x_s = y_s \quad \forall s \neq i, \quad x_i = -y_i \Rightarrow Q(x,y) = \frac{1}{|V|}\,\frac{e^{-\beta H(y)}}{e^{-\beta H(x)} + e^{-\beta H(y)}}$$

$$= \frac{1}{|V|} \cdot \frac{1}{1 + e^{\beta(H(y) - H(x))}} = \frac{1}{|V|}\,\frac{1}{1 + e^{-\beta x_i S_i}} \quad \text{con} \quad S_i = \sum_{s \text{ t.c. } s \leftrightarrow i} x_s$$