Discrete-time Markov chains

Vers. 1.0.0

Gianluca Mastrantonio

gianluca.mastrantonio@polito.it

Processo Stocastico

Processo Stocastico

Un **processo stocastico** è una collezione di variabili aleatorie, $\{X_t\}_{t \in T}$, indicizzate da un insieme T, dove ogni variabile aleatoria X_t rappresenta lo stato di un sistema casuale in un momento t.

Formalmente, un processo stocastico è definito come:

$$\{X_t : t \in T\}$$

dove:

- T è l'insieme degli indici, che spesso rappresenta il tempo.
- X_t è una variabile aleatoria per ogni $t \in T$.

In altre parole, un processo stocastico descrive l'evoluzione di un fenomeno casuale.

Esempio: Passeggiata Aleatoria

Immaginiamo una particella che si muove su una retta. Ad ogni passo t, la particella si sposta di una quantità Y_t, che può essere positiva (avanti) o negativa (indietro). La posizione della particella dopo t passi è data da X_t.

Passeggiata Aleatoria

Consideriamo un semplice esempio di processo stocastico: la **passeggiata aleatoria** (*Random Walk*).

$$X_t = X_{t-1} + Y_t, \quad t = 1, 2, 3, \ldots$$

dove:

- $X_0 = 0$ è lo stato iniziale.
- Y_t sono variabili aleatorie indipendenti e identicamente distribuite (i.i.d.).
- X_t rappresenta la posizione della passeggiata dopo t passi.

Proprietà di Markov

Definizione - Proprietà di Markov a tempo discreto

Se $\{X_t : t \in T\}$ è un processo stocastico a valori in $\mathcal{S}$ (spazio degli stati), e se per ogni $j, i, i_{n-2}, \ldots, i_0 \in \mathcal{S}$

$$f(X_t = j | X_{t-1} = i, X_{t-2} = i_{n-2}, \ldots, X_1 = i_1, X_0 = i_0) = f(X_t = j | X_{t-1} = i)$$

allora il processo ha la proprietà di Markov

Per semplicità si indica spesso

$$P(X_t = j | X_{t-1} = i) = p(i \to j) = p(i, j)$$

A livello intuitivo ci dice che la legge di un processo a un dato tempo t non dipende dal suo intero passato, ma solo dal suo passato più recente. Un processo che ha la proprietà di Markov è una Markov chain.

Proprietà di Markov

- In questa parte tratteremo solamente processi Markoviani con spazio degli stati discreto ($X_t \in \mathcal{S}$ è una variabile aleatoria discreta)
- Il processo è markoviano di ordine p se

$$f(X_t = j | X_{t-1} = i, X_{t-2} = i_{n-2}, \ldots, X_1 = i_1, X_0 = i_0) =$$

$$f(X_t = j | X_{t-1} = i, \ldots, X_{t-p} = i_{n-p})$$

- prendiamo in considerazione solo "temporally homogeneous case": $p(i,j)$ non dipende da t

Probabilità di transizione

Le $p(i,j)$ sono chiamate probabiltà di transizione, e nel caso di spazio degli stati discreto, con cardinalità di $\mathcal{S}$ pari a k, possiamo organizzare le probabilità in una matrice (di probabilità o stocastica)

$$P = \begin{pmatrix} p(1,1) & p(1,2) & \dots & p(1,k) \\ p(2,1) & p(2,2) & \dots & p(2,k) \\ \dots & \dots & \dots & \dots \\ p(k,1) & p(k,2) & \dots & p(k,k) \end{pmatrix}$$

dove abbiamo

$$\sum_{j \in \mathcal{S}} p(i,j) = 1$$

Si può facilmente generalizzare al caso in cui $\mathcal{S}$ sia infinito numerabile

La probabilità di una traiettoria

$$X_t = i_t, X_{t-1} = i_{t-1}, \dots, X_0 = i_0$$

Probabilità di transizione

si può calcolare facilmente calcolare come

$$p(i_{t-1}, i_t)p(i_{t-2}, i_{t-1}) \dots p(i_0, i_1)P(X_0 = i_0)$$

dove $\alpha(i) = P(X_0 = i_0)$ è la distribuzione iniziale, che possono essere messe in un vettore $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)$ o una matrix $P_0 = \mathrm{diag}(\boldsymbol{\alpha})$
Possiamo facilmente calcolare

$$p^{(2)}(i, j) = P(X_{t+2} = j | X_t = i) =$$

$$\sum_{k \in \mathcal{S}} P(X_{t+2} = j | X_{t+1} = k, X_t = i) P(X_{t+1} = k | X_t = i)$$

$$\sum_{k \in \mathcal{S}} p(k, j)p(i, k)$$

che può essere scritta come matrice di probabilità

$$P^{(2)} = PP = P^2$$

In generale, abbiamo che la seguente equazione di Chapman Kolmogorov

$$P^{(n+m)} = P^{(n)}P^{(m)} = P^{n+m}$$

Stati Ricorrenti

Indichiamo con T_y il tempo della prima visita allo stato y, senza contare X_0:

$$T_y = \min\{t \geq 1 : X_t = y\}$$

Se $X_0 \neq$ y, allora T_y è chiamata **hitting-time** di y, altrimenti è il **return-time**. Possiamo definire la variabile aleatoria

$$\rho_{xy} = P(T_y < \infty | X_0 = x)$$

che indica la probabilità di visitare y se si parte da x. Se

- $\rho_{yy} = 1$ allora y è uno stato ricorrente $\rightarrow$ y verrà visitato infinite volte
- $\rho_{yy} < 1$ allora y è uno stato transiente $\rightarrow$ y verrà visitato un numero finito di volte, e quindi, esiste un t' grande abbastanza, per cui per ogni $t > t'$ y non verrà più visitato.

Irriducibilità e Periodicità

Se esiste un $m \geq 0$ tale che

$$p^{(m)}(i,j) = P(X_{t+m} = j | X_t = i) > 0$$

diciamo che j è raggiungibile da i, e lo indichiamo con $i \to j$. Per definizione abbiamo che

$$p^{(0)}(i,j) = 0 \text{ se } i \neq j, \qquad p^{(0)}(i,i) = 1$$

Se

$$i \to j \qquad j \to i$$

allora gli stati i e j sono **comunicanti** e possiamo scrivere $i \leftrightarrow j$

- Una Markov chain si dice **irriducibile** se tutti gli stati sono comunicanti.
- Uno stato j si dice **periodica** di periodo $T > 1$ se $p^{(k)}(j,j) > 0$ solo per k multipli di T. Se il periodo è uno, si dice **aperiodica**
- Se una MC è irriducibile e aperiodica (tutti gli stati sono aperiodici), allora è **ergodica**

Distribuzione Stazionaria e Limite

- Una **distribuzione stazionaria** π è una distribuzione su $\mathcal{S}$ tale che se $X_0 \sim \pi$, allora $X_1 \sim \pi$, e per ogni t, $X_t \sim \pi$.
- una **distribuzione limite** è una distribuzione π tale che

$$\lim_{t \to \infty} p^{(t)}(x, y) = \pi(y)$$

Data la distribuzione iniziale $\boldsymbol{\alpha}$, possiamo calcolare la distribuzione $\boldsymbol{\alpha}_1$ su X_1 come

$$\boldsymbol{\alpha}_1 = \boldsymbol{\alpha} P$$

e in generale la distribuzione su X_t come

$$\boldsymbol{\alpha}_t = \boldsymbol{\alpha}_{t-1} P$$

Quindi se la MC ha una distribuzione stazionaria, dobbiamo avere che

$$\pi = \pi P \Rightarrow \pi(j) = \sum_{i \in \mathcal{S}} \pi(i) p(i, j)$$

Distribuzione Stazionaria e Limite

Importante: Se $\nu(x)$ è una distribuzione limite, segue che è anche distribuzione stazionaria, ma l'altro verso dell'implicazione non è vero.
Prendete per esempio una transizione

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

e si può verificare che la distribuzione $(0.5, 0.5)$ è la distribuzione stazionaria

Theorem

Se la catena è irriducibile, aperiodica, e esiste una distributione stazionaria π, allora per ogni $i, j \in \mathcal{S}$

$$\lim_{t \to \infty} P(X_t = j | X_0 = i) = \pi(j)$$

quindi π è la distribuzione limite

Detailed Balance condition

Una distribuzione π, con una transizione $p()$ è detailed balanced se per ogni $x, y \in \mathcal{S}$ abbiamo

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

Se sommiamo rispetto a y abbiamo che

$$\sum_{y \in \mathcal{S}} \pi(y)p(y,x) = \pi(x) \sum_{y \in \mathcal{S}} p(x,y) = \pi(x)$$

e quindi

$$\pi = \pi P$$

e π è una distribuzione stazionaria

I risultati delle ultime due slides valgono anche nei casi in cui lo spazio degli stati non sia discreto.

Markov chain e Statistica Bayesiana

Markov Chain per statistica Bayesiana

Nella statistica Basyesiana avremo una distribuzione, chiamata **a-posteriori** da cui vogliamo simulare (Monte Carlo). Per farlo creeremo una Markov change

- dove la distribuzioni stazionaria è la a-posteriori
- la transizione rispetta la condizione di detailed balance
- la catena è irriducibile e aperiodica

Facendo correre l'algoritmo per abbastanza iterazioni, arriveremo alla distribuzione limite (che non dipende più da dove la catena è partita) che è la a-posteriori. L'algoritmo si chiama Markov chain Monte Carlo (MCMC).

Per la dimostrazione che l'MCMC effettivamente funzione in tutte le situazioni, abbiamo bisogno di risultati più generali e complessi, visto che lavoriamo con variabili miste, che però vanno oltre lo scopo del corso.