IC - Cadeias de Markov

Victor Moreli dos Santos

Universidade de São Paulo

victormoreli@usp.br

May 31, 2024

Overview

Processos Estocásticos

Processo Estocástico

Um processo estocástico é um conjunto de variáveis aleatórias $\{X_t\}_{t=0}^{\infty}$ onde cada X_t tem o mesmo range $\{S_1, S_2, \ldots, S_n\}$. Esse range é chamado de espaço de estados.

Na prática, se trata de um processo que é modelado por estados e as transições entre eles. Quando o tempo é considerado como discreto, o número de estados é finito.

Cadeias de Markov

Cadeia de Markov

Uma cadeia de Markov é um processo estocástico que satisfaz a propriedade de Markov:
A probabilidade do estado seguinte depende apenas do estado atual, e não dos anteriores.

$$P(X_{t+1} = S_j | X_t = S_{i_t}, X_{t-1} = S_{i_{t-1}}, \dots, X_0 = S_{i_0})$$
$$= P(X_{t+1} = S_j | X_t = S_{i_t})$$

São chamados de processos sem memória, devido a essa propriedade.

Cadeias de Markov

Matriz Estocástica

A matriz estocástica é uma matriz não negativa onde a soma de cada linha é igual a 1.

Definindo-se a probabilidade de transição do estado i para o j como:

$$p_{ij}(t) = P(X_t = S_j | X_{t-1} = S_i)$$

Matriz de Transição

Pode-se montar uma matriz de transição P_{nxn} tal que:

$$P_{nxn}(t) = [p_{ij}(t)]$$

Se as probabilidades de transição não variam com o tempo, a cadeia é dita estacionária/homogênea e a matriz de transição é uma matriz estocástica constante:

$$P = [p_{ij}]$$

Toda matriz estocástica P_{nxn} define uma Cadeia de Markov de n estados, pois as entradas da matriz podem ser interpretadas como probabilidades de transição. A recíproca também é verdadeira.
Também modela-se uma matriz estocástica por meio de um grafo orientado e valorado:

- nós - estados
- arestas - transições
  - orientação - origem e destino
  - peso - probabilidade de transição

Cadeias de Markov

Vetor de Probabilidade / Vetor Estocástico

Um vetor de probabilidade é dado por:

$$p^T(k) = (p_1(k), p_2(k), \ldots, p_n(k)) \text{ onde } p_j(k) = P(X_k = S_j) \text{ e } \sum_k p_k = 1$$

Ou seja, seus elementos são as probabilidades de estar em cada estado após k passos. Nota-se que todas as linhas de matrizes estocásticas são vetores desse tipo. Um vetor estocástico importante é o da distribuição inicial:

$p^t(0)$: Define probabilidades de cada estado ser o estado inicial.

Objetivos

- Descrever $p^t(k)$ para todo $p^t(0)$
- Determinar a existência e, se possível, o valor de

$$lim_{k \to \infty} p^t(k)$$

- Se o limite acima não existir, determinar a possibilidade de haver um limite de Cesáro e seu significado:

$$\lim_{k \to \infty} \left[\frac{p^t(0) + p^t(1) + \cdots + p^t(k-1)}{k} \right]$$

Matriz de Transição de k passos

Matriz de Transição de k passos

Pode-se mostrar que:

$$p^t(k) = p^t(0)P^k$$

O que significa que o vetor de probabilidade de k passos é dado pelo produto da distribuição inicial com a matriz de transição elevada a k.

A probabilidade de ir do estado i ao j em k passos é dada pela própria entrada (i, j) da matriz P^k.

Redutibilidade e Grafos

Matriz Redutível

A é uma matriz redutível se existir uma matriz de permutação P tal que: $P^TAP = \begin{bmatrix} X & Y \\ 0 & D \end{bmatrix}$, sendo X e Z ambas quadradas. Senão, A é dita irredutível.
P^TAP é chamada de permutação simétrica de A, o que significa mudar linhas e colunas da mesma forma.

Para transpor esses conceitos para grafos, primeiramente define-se o grafo dirigido e valorado com base na matriz de adjacência A como $G(A)$. Quando P é uma matriz de permutação, $G(P^TAP) = G(A)$, o efeito é simplesmente uma troca nos nomes dos nós.

Grafo fortemente conexo

Uma matriz A é irredutível se, e somente se, G(A) é fortemente conexo.

Teorema de Perron-Frobenius

Teorema de Perron-Frobenius

Sendo A_{nxn} não negativa e irredutível, então:

- A tem um autovalor simples (de multiplicidade algébrica 1) r, que também é o seu raio espectral: $r = \rho(A)$ e $r > 0$
- Existe um autovetor $x > 0$ correspondente a r: $Ax = rx$
- Existe um vetor único definido por:

  $$Ap = rp,\ p > 0, \text{ e } \|p\| = 1$$

  p é o vetor de Perron de A. Não há outros autovetores não negativos para A (independente do autovalor), exceto por múltiplos positivos de p.
- Fórmula de Collatz-Wielandt $r = \max_{x \in N} f(x)$,

  $$f(x) = \min_{1 \leq i \leq n} \frac{[Ax]_i}{x_i}, \text{ com } N = \{x \mid x \geq 0, x \neq 0\}$$

  que é basicamente uma forma de encontrar o raio espectral de A sobre todos os vetores não nulos com componentes não negativas.

Primitividade

Primitividade de Matrizes

Uma matriz A irredutível e não negativa é dita:

- Primitiva se A tem somente 1 autovalor
- Imprimitiva se A tem $h > 1$ autovalores, sendo h o índice de imprimitividade da matriz

A é primitiva se, e somente se:

$$\lim_{k\to\infty} \left(\frac{A}{r}\right)^k = G = pq^T(q^Tp)^{-1} > 0$$

onde p e q são os vetores de Perron para A e A^T respectivamente, e $r = \rho(A)$ (r é o único autovalor de A, que é o raio espectral).

Voltando às cadeias de Markov

Dividindo as cadeias de Markov em 4 categorias mutualmente exclusivas:

- (I) Irredutível com $lim_{k \to \infty} P^k$ existente
- (II) Irredutível com $lim_{k \to \infty} P^k$ não existente
- (III) Redutível com $lim_{k \to \infty} P^k$ existente
- (IV) Redutível com $lim_{k \to \infty} P^k$ não existente

Cadeias de Markov Irredutíveis

No caso (I), P é primitiva, pois o limite existe e sabemos exatamente o quanto ele vale. Sendo e/n (vetor de distribuição uniforme) o vetor de Perron de P e π o vetor de Perron de P^T:

$$\lim_{k\to\infty} P^k = \frac{(e/n)\pi^t}{\pi^t(e/n)} = \frac{e\pi^T}{\pi^T e} = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \cdots & \pi_n \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix} > 0$$

Logo, uma distribuição de probabilidade limitante existe e é dada por:

$$\lim_{k\to\infty} p^T(k) = \lim_{k\to\infty} p^T(0)P^k = p^T(0)e\pi^T = \pi^T$$

Conclusão

Quando P (matriz estocástica que define a transição de estados) é primitiva e irredutível, a distribuição de probabilidade limitante é igual ao vetor de Perron de P^T.
Nota-se que essa probabilidade limitante independe da distribuição inicial.

Cadeias de Markov Irredutíveis

No caso (II), P é imprimitiva e portanto o limite não existe. Mas o limite de Cesàro existe e tem a mesma interpretação do limite anterior:

$$\lim_{k\to\infty} \frac{I+P+\cdots+P^{k-1}}{k} = \frac{e}{n}\pi^T\pi^T\left(\frac{e}{n}\right) = e\pi^T\pi^T e = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_n \\ \pi_1 & \pi_2 & \cdots & \pi_n \\ \vdots & \vdots & \ddots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_n \end{pmatrix}$$

O que é similar ao caso de P primitiva, logo:

$$\lim_{k\to\infty} \frac{p^T(0)+p^T(1)+\cdots+p^T(k-1)}{k} = \lim_{k\to\infty} p^T(0)\frac{I+P+\cdots+P^{k-1}}{k} = p^T(0)e\pi^T = \pi^T$$

A interpretação é a mesma: o j-ésimo componente de π^T representa a fração de tempo que a cadeia fica no estado j pensando em $t \to \infty$.

Cadeias de Markov Irredutíveis - Resumo

Cadeias de Markov Irredutíveis

Seja P a matriz de probabilidade de transição para uma cadeia de Markov irreduzível com estados $\{S_1, S_2, \ldots, S_n\}$, e seja π^T o vetor de Perron de P^T. É valido que para toda distribuição inicial $p^T(0)$:

- A matriz de transição de k-ésimo passo é P^k porque a entrada (i, j) em P^k é a probabilidade de mover-se de S_i para S_j exatamente em k passos.
- O vetor de distribuição de k-ésimo passo é dado por $p^T(k) = p^T(0)P^k$.
- Se P é primitiva, e se e denota a coluna de todos os 1's, então $\lim_{k\to\infty} P^k = e^{\pi^T}$ e $\lim_{k\to\infty} p^T(k) = \pi^T$.
- Se P é imprimitiva, então $\lim_{k\to\infty} \frac{I+P+\cdots+P^{k-1}}{k} = e^{\pi^T}$ e $\lim_{k\to\infty} \frac{p^T(0)+p^T(1)+\cdots+p^T(k-1)}{k} = \pi^T$.

Cadeias de Markov Irredutíveis - Resumo

Cadeias de Markov Irredutíveis

- Independentemente de P ser primitiva ou imprimitiva, o componente π_j de π^T representa a fração de tempo de longo prazo que a cadeia passa no estado S_j.
- π^T é frequentemente chamado de vetor de distribuição estacionária para a cadeia, pois é o vetor de distribuição único que satisfaz $\pi^T P = \pi^T$.

Cadeias de Markov Redutíveis

Como o teorema de Perron Frobenius não é diretamente aplicável a cadeias redutíveis, a ideia é tentar aproximar a matriz de uma irredutível o máximo possível, aplicando as sucessivas permutações simétricas:

$$Q^T P Q = \begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix}$$

$$\begin{pmatrix} X & Y \\ 0 & Z \end{pmatrix} \sim \begin{pmatrix} R & S & T \\ 0 & U & V \\ 0 & 0 & W \end{pmatrix}$$

Até chegar na chamada forma canônica para matrizes redutíveis.

Cadeias de Markov Redutíveis

Forma Canônica para matrizes redutíveis

Sendo P uma cadeia de Markov redutível, sua forma canônica é dada por:

$$P \sim \begin{pmatrix} P_{11} & P_{12} & \cdots & P_{1r} & P_{1,r+1} & P_{1,r+2} & \cdots & P_{1m} \\ 0 & P_{22} & \cdots & P_{2r} & P_{2,r+1} & P_{2,r+2} & \cdots & P_{2m} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_{rr} & P_{r,r+1} & P_{r,r+2} & \cdots & P_{rm} \\ 0 & 0 & \cdots & 0 & P_{r+1,r+1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & P_{r+2,r+2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & P_{mm} \end{pmatrix} \text{ onde:}$$

- P_{kk} com $1 \leq k \leq r$ são as classes transientes, às quais não é possível retornar após sair delas
- $P_{r+j,r+j}$ com $j \geq 1$ são as classes ergódicas, também chamadas de recorrentes, as quais são cadeias irredutíveis que compõem a cadeia redutível maior. Assim que alcançadas, o sistema não sai delas.

Fim