

Notas Processos estocásticos
Nestor Caticha
23 de abril de 2012

O Teorema de Perron Frobenius para matrizes de Markov

Considere um processo estocástico representado por um conjunto de variáveis aleatórias $\{X_i\}$ onde o índice i pode ser considerado como um tempo discreto. O valor da variável X_i é x_i que toma valores num conjunto $L = \{\alpha_1, \alpha_2, \dots, \alpha_K\}$. A probabilidade do evento $\hat{X}_{0,n} = \{X(t_n) = x(t_n), X(t_{n-1}) = x(t_{n-1}), \dots, X(t_0) = x(t_0)\}$ é denotada por $\mathbb{P}(\hat{X}_{0,n})$. A regra do produto para essa sequência leva a

$$\mathbb{P}(\hat{X}_{0,n}) = \mathbb{P}(X_n = x_n | \hat{X}_{0,n-1}) \mathbb{P}(\hat{X}_{0,n-1})$$

Para o caso Markoviano em que a única informação relevante é o último valor de x

$$\mathbb{P}(\hat{X}_{0,n}) = \mathbb{P}(x_n | x_{n-1}) \mathbb{P}(\hat{X}_{0,n-1})$$

que pode ser estendido a

$$\mathbb{P}(\hat{X}_{0,n}) = \prod_{i=1,n} \mathbb{P}(x_i | x_{i-1}) \mathbb{P}(x_0)$$

Seja $\mathbb{M}$ a matriz de transição de Markov. É uma matriz quadrada $K \times K$ com elementos não negativos:

$$M_{ij} = \mathbb{P}(x_n = \alpha_j | x_{n-1} = \alpha_i),$$

é a probabilidade de transição (1-passo) do estado i para o j . Consideramos o caso em que estes elementos de matriz não dependem do tempo. Note que $\sum_j M_{ij} = 1$, mas $\sum_i M_{ij}$ não é obrigatoriamente 1.

A probabilidade $P(x_n)$ é obtida marginalizando sobre todas as variáveis X_i , $i = 0, \dots, n-1$, em notação de matriz

$$P_n = \Pi_0 \mathbb{M}^n \tag{1}$$

onde P_n e Π_0 , vetores (linha) de dimensão K , são respectivamente as probabilidades no instante n e 0.

Mostraremos a seguir vários resultados que coletados são um caso particular do teorema de Perron-Frobenius:

1. $\mathbb{M}$ tem um autovetor à direita v_1 que é o vetor coluna com todas as entradas iguais a 1.
2. O autovalor associado a esse autovetor é $\lambda^{PF} = 1$ de multiplicidade algébrica e geométrica 1
3. Todos os outros autovalores λ_i de $\mathbb{M}$ satisfazem $|\lambda_i| < 1$
4. O autovetor à esquerda u^{PF} associado a λ^{PF} tem todas as componentes não nulas. Pode ser normalizado de forma a que a soma das componentes seja 1.

5. Para cada um dos outros autovetores de $\mathbf{M}$ à esquerda ou à direita, a soma das componentes é zero.
6. O vetor P_n tende exponencialmente rápido com n para u^{PF}

Prova de 1 e 2: Como os elementos de $\mathbf{M}$ são probabilidades temos $\sum_j M_{ij} = 1$ lembrando que $v^1 = (1, \dots, 1)^T$ pode ser escrito

$$\sum_j M_{ij}(v^1)_j = (v^1)_i$$

ou $\mathbf{M}v^1 = v^1$

Prova de 3. Seja $v = (v_1, v_2, \dots, v_K)^T$ um autovetor com autovalor associado λ

$$\lambda v_i = \sum_j M_{ij}v_j,$$

existe uma componente que satisfaz $|v_k| \geq |v_i|$ para todo $i \neq k$. Tomando o módulo da equação de autovalores, temos uma primeira desigualdade

$$|\lambda v_i| = \left| \sum_j M_{ij}v_j \right| \leq \sum_j M_{ij}|v_j| \quad (2)$$

e uma segunda desigualdade é obtida majorando $|v_j|$ por $|v_k|$

$$|\lambda v_i| = \left| \sum_j M_{ij}v_j \right| \leq \sum_j M_{ij}|v_j| \leq \sum_j M_{ij}|v_k| = |v_k| \quad (3)$$

pois a somatória é 1 por normalização das probabilidades. Assim temos que, para todo i

$$|\lambda v_i| \leq |v_k| \quad (4)$$

e em particular para $i = k$

$$|\lambda||v_k| \leq |v_k| \quad (5)$$

Dois casos são possíveis. Se $|\lambda| = 1$, então as duas desigualdades acima são igualdades. A desigualdade 2 mostra que todos os termos v_i tem a mesma fase. A segunda desigualdade em 3 mostra que os v_i também tem o mesmo módulo. A conclusão é que se $|\lambda| = 1$ então o vetor v so pode ser um múltiplo de $(1, \dots, 1)^T$ e que $\lambda = 1$, é portanto o único o autovalor 1, i.e. simples. Note que vale o inverso: se v for um múltiplo de v^1 implica $|\lambda| = 1$.

O segundo caso ocorre se v não for um múltiplo de $(1, \dots, 1)^T$, então não pode valer a igualdade: $|\lambda| < 1$.

Prova de 4: que o autovetor à esquerda u^{PF} de autovalor 1 tem as componentes positivas. Para qualquer autovetor à esquerda temos

$$\lambda u_j = \sum_i u_i M_{ij} \quad (6)$$

tomando valor absoluto dos dois lados

$$|\lambda||u_j| = \left| \sum_i u_i M_{ij} \right| \quad (7)$$

a desigualdade triangular

$$|\lambda||u_j| = \left| \sum_i u_i M_{ij} \right| \leq \sum_i |u_i| M_{ij} \quad (8)$$

somando sobre todo os j

$$|\lambda| \sum_j |u_j| \leq \sum_i |u_i| \sum_j M_{ij} = \sum_i |u_i| \quad (9)$$

Se $|\lambda| < 1$ não diz nada, mas se $\lambda = 1$ significa que a desigualdade acima é uma igualdade. portanto a expressão 8 é uma igualdade e todas as componentes tem a mesma fase, que podemos tomar igual a zero. Segue que todas as componentes são positivas.

Prova de 5. Consideremos outro autovetor à esquerda u com autovalor $|\lambda| < 1$.

$$\lambda u_j = \sum_i u_i M_{ij} \quad (10)$$

$$\lambda \sum_j u_j = \sum_{ij} u_i M_{ij} = \sum_i u_i \sum_j M_{ij} \quad (11)$$

$$\lambda \sum_j u_j = \sum_i u_i \quad (12)$$

e portanto para qualquer autovetor à esquerda com autovalor menor que 1, a soma das componentes deve ser zero.

Prova de 6: Convergência para o equilíbrio. Denote, para $a = 2, \dots, K$, os autovalores menores que 1 por $\{\lambda_a\}$ e por $\{u^a\}$ os autovetores associados à esquerda. Note que acima provamos

$$\sum_i u_i^{PF} = 1, \quad \sum_i u_i^a = 0; \quad (13)$$

A condição inicial Π_0 pode ser escrita na base dos autovetores à esquerda

$$\Pi_0 = c_0 u^{PF} + \sum_{a=2}^K c_a u^a \quad (14)$$

Somando as componentes dos vetores acima, por normalização temos que $1 = c_0 \times 1 + 0$ ou seja

$$\Pi_0 = u^{PF} + \sum_{a=2}^K c_a u^a \quad (15)$$

e usando a equação (1)

$$P_n = \Pi_0 \mathbf{M}^n = u^{PF} + \sum_{a=2}^K c_a \lambda_a^n u^a \quad (16)$$

e podemos mostrar a convergência em qualquer norma apropriada

$$\|P_n - u^{PF}\| = \left\| \sum_{a=2}^K c_a \lambda_a^n u^a \right\| \quad (17)$$

ordene os autovetores de forma que $|\lambda_2| \geq |\lambda_a|$ para todos os $a > 2$:

$$\|P_n - u^{PF}\| = |\lambda_2|^n \left\| \sum_{a=2}^K c_a \left(\frac{\lambda_a}{\lambda_2}\right)^n u^a \right\| < C e^{-n/\tau} \rightarrow 0 \quad (18)$$

onde $\tau = 1/\ln|\lambda_2|^{-1}$ é o tempo característico de termalização.

Cadeias de Markov Absorventes

O material desta seção foi extraído de Grinstead e Snell ¹ que recomendamos como fonte de consulta.

¹http://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/book.html

Há processos markovianos que ao chegar a um determinado estado cessam de evoluir. Isto nos leva à :

Definição: Estado Absorvente (EA): Se para todo $j \neq i$, $M_{ij} = 0$, então o estado i é um *estado absorvente*.

Definição: Cadeias de Markov Absorventes (CMA) são as cadeias de Markov com ao menos um estado absorvente.

Definição: Estado Transiente (ET) é todo estado de uma CMA que não é absorvente.

Estamos interessados em determinar propriedades do seguinte tipo: qual é a esperança do tempo que leva para chegar a um estado absorvente particular, a qualquer estado absorvente, qual é a esperança do número de visitas a um estado transiente antes de ser absorvido. Considere uma CMA com n_t estados transientes e n_r estados absorventes, $n_t + n_r = K$. A matriz de Markov da CMA pode ser escrita assim

$$\mathbf{M} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}$$

onde $\mathbf{Q}$ é uma matriz $n_t \times n_t$ que descreve as transições entre estados transientes, $\mathbf{R}$ é a matriz $n_t \times n_r$ que descreve a transição de estados transientes a absorventes. A matriz $\mathbf{0}$ é $n_r \times n_t$ e seu elementos são todos nulos. A matriz $\mathbf{1}$ é diagonal, $(\mathbf{1})_{ij} = \delta_{ij}$ e tem dimensão $n_r \times n_r$. A equação (1) continua descrevendo a dinâmica e precisamos encontrar as potências de $\mathbf{M}$:

$$\mathbf{M}^n = \begin{pmatrix} \mathbf{Q}^n & f(\mathbf{R}, \mathbf{Q}) \\ \mathbf{0} & \mathbf{1} \end{pmatrix}$$

Devido à absorção esperamos que $\mathbf{Q}^n \rightarrow \mathbf{0}$ pois $\mathbf{Q}$ descreve as transições entre os estados transientes e há a possibilidade de

transição de algum estado transiente para um estado absorvente: a probabilidade de estar em um estado transiente vai a zero quando o tempo n cresce.

Há várias questões da dinâmica que podem ser respondidas a partir de $\mathbf{Q}$, em particular olharemos para algumas delas. Sejam i e j dois estados transientes e l um estado absorvente:

- Dado que o sistema está no estado i qual é o valor esperado do número de visitas ao estado j antes de ser absorvido?
- Começando do estado i , qual é o valor esperado do tempo que demora o sistema para ser absorvido ?
- Começando do estado i , qual é a probabilidade que o sistema seja absorvido no estado l .

Começamos definindo a matriz chamada fundamental $\mathbf{N}$ como o inverso de $(\mathbf{1} - \mathbf{Q})$:

$$(\mathbf{1} - \mathbf{Q})(\mathbf{1} + \mathbf{Q} + \mathbf{Q}^2 + \dots + \mathbf{Q}^n) = \mathbf{1} - \mathbf{Q}^{n+1}.$$

Para $n \rightarrow \infty$,

$$\mathbf{N} = \lim_{n \rightarrow \infty} \sum_{i=0}^n \mathbf{Q}^i, \quad (19)$$

temos que $(\mathbf{1} - \mathbf{Q})\mathbf{N} = \mathbf{1}$.

Para i e j transientes (fixos) e l um número de passos da dinâmica, definimos a variável estocástica $Y^{(l)}$ que toma valor 1 se, começando do estado i , depois de l passos, estiver no estado j , e zero se não estiver. Então a probabilidade

$$P(Y^{(l)} = 1) = 1 - P(Y^{(l)} = 0) = q_{ij}^{(l)}$$

Exercício Convença-se que o valor esperado de $Y^{(l)}$ é

$$E[Y^{(l)}] = q_{ij}^{(l)}$$

é o valor esperado de estar em j após l passos. Somando sobre todos os valores de l , temos o valor esperado de visitação de j antes de ser absorvido:

$$E\left[\sum_{l=0}^{\infty} Y^{(l)}\right] = \sum_{l=0}^{\infty} q_{ij}^{(l)}$$

que pela definição de $\mathbf{N}$, é

$$E\left[\sum_{l=0}^{\infty} Y^{(l)}\right] = n_{ij}$$

O tempo médio até a absorção depende do estado inicial i , t_i . É a soma dos tempos esperados gastos em cada um dos estados transientes, portanto

$$t_i = \sum_j n_{ij}$$

Consideramos agora um estado absorvente a e um estado inicial tranisente i . Queremos saber qual é a probabilidade de a ser o estado em que o processo para de evoluir, que chamaremos B_{ia} . Consideremos um outro estado absorvente j , e a probabilidade de depois de l passos o sistema não ter sido absorvido e esteja em j . A regra do produto pode ser aplicada a probabilidade de: sair de i , chegar a j após l passos e ser absorvida em a :

$$P(i \rightarrow^l j, j \rightarrow a) = q_{ij}^{(l)} r_{ja}.$$

A regra da soma, para eventos mutuamente excludentes, nos dá

$$B_{ia} = \sum_l \sum_j q_{ij}^{(l)} r_{ja},$$

isto é, somamos sobre todos os estados transientes j que podem levar ao estado absorvente a em $l + 1$ passos, e posteriormente somamos sobre todos os valores de l . Mudando a ordem das somas e usando a expressão 19 obtemos

$$B_{ia} = \sum_j \sum_l q_{ij}^{(l)} r_{ja} = \sum_j n_{ij} r_{ja}$$

ou $\mathbb{B} = \mathbb{N}\mathbb{R}$.