

Demonstrações dos resultados de convergência e estacionariedade para cadeias finitas (com finitos estados).

1 Convergência $P^n \rightarrow \Pi$ para $P > 0$

Vamos demonstrar de modo elementar

Teorema 1. *Se P é matriz estocástica positiva, isto é $p_{ij} > 0$ ($\forall i, j$), então $\lim_{n \rightarrow \infty} P^n$ existe e é uma matriz estocástica com linhas iguais a um vetor-linha invariante π*

$$\lim_{n \rightarrow \infty} P^n = \mathbf{1}\pi$$

onde $\mathbf{1}$ é o vetor-coluna de uns e π é a única distribuição invariante (posto geométrico 1).

Para a demonstração usaremos o seguinte resultado:

Proposição 2. *Seja $P = (p_{i,j})$ uma matriz estocástica com todags as entradas positivas e defina*

$$\varepsilon := \min_{i,j} \{ p_{i,j} \} > 0.$$

Seja $\mathbf{c} = (c_i)$ um vetor coluna qualquer e defina

$$\begin{aligned} m_0 &= \min_i c_i, & M_0 &= \max_i c_i, \\ m_1 &= \min P\mathbf{c}, & M_1 &= \max P\mathbf{c}. \end{aligned}$$

Então valem

$$m_1 > m_0, \quad M_1 < M_0, \quad e \quad M_1 - m_1 \leq (1 - 2\varepsilon)(M_0 - m_0).$$

Demonstração. Se $M_0 = m_0$ então $\mathbf{c}$ é constante $P\mathbf{c} = \mathbf{c}$, de modo que $m_1 = m_0$, $M_1 = M_0$ e todas as desigualdades desejadas são triviais. Assim suponha $M_0 > m_0$.

Escolha um índice j_0 tal que $c_{j_0} = M_0$ e um índice j_1 tal que $c_{j_1} = m_0$. Defina os vetores auxiliares

$$\mathbf{c}' := (c'_i), \quad c'_i = \begin{cases} M_0, & i = j_0, \\ m_0, & i \neq j_0, \end{cases} \quad \mathbf{c}'' := (c''_i), \quad c''_i = \begin{cases} m_0, & i = j_1, \\ M_0, & i \neq j_1. \end{cases}$$

Pelo modo como foram construídos, tem-se, coordenada a coordenada,

$$c' \leq c \leq c''.$$

Como P tem entradas não-negativas, a multiplicação por P preserva a ordem coordenada a coordenada, logo

$$Pc' \leq Pc \leq Pc''.$$

Para i fixo, calculamos

$$(Pc')_i = \sum_j p_{i,j} c'_j = m_0 \sum_{j \neq j_0} p_{i,j} + M_0 p_{i,j_0} = m_0 + p_{i,j_0} (M_0 - m_0).$$

Como $p_{i,j_0} \geq \varepsilon$ para todo i , obtemos

$$(Pc')_i \geq m_0 + \varepsilon(M_0 - m_0),$$

para todo i . Portanto,

$$m_1 = \min_i (Pc)_i \geq \min_i (Pc')_i \geq m_0 + \varepsilon(M_0 - m_0).$$

Analogamente,

$$(Pc'')_i = M_0 - p_{i,j_1} (M_0 - m_0) \leq M_0 - \varepsilon(M_0 - m_0),$$

para todo i , e daí

$$M_1 = \max_i (Pc)_i \leq \max_i (Pc'')_i \leq M_0 - \varepsilon(M_0 - m_0).$$

Subtraindo as duas estimativas, segue

$$M_1 - m_1 \leq (M_0 - \varepsilon(M_0 - m_0)) - (m_0 + \varepsilon(M_0 - m_0)) = (1 - 2\varepsilon)(M_0 - m_0),$$

concluindo a prova. □

Demonstração do teorema Seja $\varepsilon = \min_{i,j} p_{ij} > 0$. Fixe uma coluna $j \in \{1, \dots, n\}$ e ponha $\mathbf{e}_j$ o vetor canônico. Defina

$$a_k := \min_i (P^k \mathbf{e}_j)_i, \quad b_k := \max_i (P^k \mathbf{e}_j)_i.$$

Pela proposição anterior temos, para todo k ,

$$a_{k+1} \geq a_k, \quad b_{k+1} \leq b_k, \quad b_{k+1} - a_{k+1} \leq (1 - 2\varepsilon)(b_k - a_k).$$

Como $b_0 - a_0 = 1$ segue $b_k - a_k \leq (1 - 2\varepsilon)^k \rightarrow 0$, e portanto a_k e b_k convergem para o mesmo limite π_j . Para $k \geq 1$, a positividade de P implica $(P^k \mathbf{e}_j)_i > 0$ para todo i , logo $0 < \pi_j < 1$.

Isso mostra que a j -ésima coluna de P^k converge para o vetor constante $\pi_j \mathbf{1}$. Como isso vale para todo j , existe a matriz limite

$$\lim_{k \rightarrow \infty} P^k = Q = \mathbf{1} \pi,$$

onde $\pi = (\pi_1, \dots, \pi_n)$ é a linha formada pelos limites das colunas. Cada linha de Q soma 1 (passando ao limite das linhas de P^k), portanto $\sum_j \pi_j = 1$.

Passando ao limite em $P^{k+1} = P^k P$ obtemos $Q = QP$. Escrevendo $Q = \mathbf{1} \pi$ obtemos $\mathbf{1}(\pi P) = \mathbf{1} \pi$, isto é, $\mathbf{1}(\pi P - \pi) = 0$, o que implica $\pi P = \pi$. Assim π é uma distribuição invariante. Se ν fosse outra distribuição invariante, então $\nu = \nu P^k \rightarrow \nu Q = \nu(\mathbf{1} \pi) = (\nu \mathbf{1}) \pi = \pi$, logo $\nu = \pi$ e a invariante é única. Isto conclui a prova. $\square$

Lema 3. *Seja P a matriz de transição de um passeio aleatório aperiódico. Existe um natural t_0 tal que para todo i , vale $p_{i,i}^{(t)} > 0$ para todo $t \geq t_0$.*

Demonstração. Observemos que para cada vértice i o conjunto $\mathcal{T}(i)$ é fechado para a adição pois, como vimos acima, na prova da proposição, se $a, b \in \mathcal{T}(i)$ então $p_{i,i}^{(a+b)} > 0$.

Sejam $m_1, m_2, \dots, m_k \in R_i$ tais que $\text{mdc}(m_1, m_2, \dots, m_k) = 1$, os quais existem pela hipótese de aperiodicidade e definição de mdc. Pelo teorema de Bézout (apêndice ??) existem inteiros positivos $x_1, \dots, x_{k+}, y_1, \dots, y_{k-}$, onde $k^+ + k^- = k$ e tais que

$$1 = (m_1 x_1 + m_2 x_2 + \dots + m_{k+} x_{k+}) - (m_1 y_1 + m_2 y_2 + \dots + m_{k-} y_{k-})$$

que reescrevemos como $1 = M - N$ onde $M = \sum_j m_j x_j$ e $N = \sum_j m_j y_j$.

Pelo fechamento para a soma em $\mathcal{T}(i)$, temos $M, N \in \mathcal{T}(i)$.

Consideremos um inteiro m grande, $m \geq NM$ basta. A divisão inteira de m por N deixa

resto $r \in \{0, 1, \dots, N-1\}$, donde $m = qN + r$. Ainda,

$$0, M, 2M, 3M, \dots, (N-1)M$$

é uma sequência de N números cujos restos da divisão por N devem ser todos distintos, portanto, um deles deixa resto r , ou seja, $aM = q'N + r$ para algum $a \in \{0, 1, 2, \dots, N-1\}$. De $aM = q'N + r$ tiramos $r = aM - q'N$ e substituindo na divisão de m por N , obtemos

$$m = (q - q')N + aM.$$

Como $M, N \in \mathcal{T}(i)$, por fechamento para a soma $(q - q')N, aM \in \mathcal{T}(i)$ e $(q - q')N + aM \in \mathcal{T}(i)$. Note que de $m > aM$ (pois $a \in \{0, \dots, N-1\}$) temos $q > q'$ de modo que $q - q' > 0$. Com isso concluímos que todo $m \in \mathbb{Z}_+$ suficientemente grande pertence a $\mathcal{T}(i)$.

Podemos, então, garantir que para todo vértice i , existe $m(i) \in \mathbb{Z}_+$ tal que $m \in \mathcal{T}(i)$ para todo $m \geq m(i)$. Para fechar a prova do lema basta tomar $t_0 = \max_i \{m(i)\}$, pois o conjunto de vértices é finito. $\square$

Corolário 4. *Seja P a matriz de transição de uma passeio aleatório irreduzível e aperiódico. Existe um inteiro positivo t_0 tal que para todos i, j vale $p_{i,j}^{(t)} > 0$ para todo $t \geq t_0$.*

Demonstração. Do lema acima, podemos garantir que para todo vértice v , existe $m(v) \in \mathbb{Z}_+$ tal que $m \in \mathcal{T}(v)$ para todo $m \geq m(v)$.

Sejam $j \neq i$ vértices e $m(i, j) \in \mathbb{Z}_+$ tal que $p_{i,j}^{(m(i,j))} > 0$, que existe pela irreduzibilidade. Fazendo $M(i, j) := m(i) + m(i, j)$ temos para todo $t \geq M(i, j)$

$$p_{i,j}^{(t)} \geq p_{i,i}^{(t-m(i,j))} p_{i,j}^{(m(i,j))} > 0$$

pois $t - m(i, j) \geq m(i)$. Finalmente, basta tomar $t_0 = \max_i \{m(i) + \max_j M(i, j)\}$. $\square$

Corolário 5. *Para toda cadeia de Markov $(X_n)_n$ finita, irreduzível e aperiódica, com matriz de transição P e distribuição inicial λ , a distribuição λP^n converge para uma única distribuição estacionária.*

Demonstração. Seja $k \in \mathbb{N}$ tal que $P^k > 0$.

$$\lim_{n \rightarrow \infty} P^n = \lim_{n \rightarrow \infty} P^{k+n} = \mathbf{1}\pi.$$

portanto

$$\lim_{n \rightarrow \infty} \lambda P^n = \lambda \mathbf{1}\pi = \pi.$$

□

2 Fórmula de Kac para cadeias finitas

Abaixo $\mathbb{E}_i[T_j] = \mathbb{E}[T_j \mid X_0 = i]$.

Teorema 6. *Seja $(X_n)_{n \geq 0}$ uma cadeia de Markov irredutível em um espaço de estados finito $S = \{1, 2, \dots, n\}$, com matriz de transição $P = (p_{ij})$ e vetor invariante $\pi = (\pi_1, \dots, \pi_n)$.*

Para cada estado $j \in S$, sejam

$$T_j^+ = \inf\{n \geq 1 : X_n = j\} \quad T_j = \inf\{n \geq 0 : X_n = j\}$$

tempos de primeiro retorno a j . Então o tempo médio de retorno a j satisfaz

$$\mathbb{E}_j[T_j^+] = \frac{1}{\pi_j}.$$

Demonstração. Seja $P = (p_{i,k})_{i,k \in S}$ a matriz de transição e $\pi = (\pi_i)_{i \in S}$ uma distribuição invariante (assumimos $\pi_j > 0$; vale automaticamente se a cadeia for finita e irredutível). Para $i \neq j$, condicionando no primeiro passo

$$m_{i,j} = \mathbb{E}_i[T_j] = 1 + \sum_{k \in S} p_{i,k} m_{k,j}.$$

Para $i = j$, usamos T_j^+ (primeiro retorno) e obtemos, de modo análogo,

$$m_j = \mathbb{E}_j[T_j^+] = 1 + \sum_{k \in S} p_{j,k} m_{k,j}.$$

Unificando ambas as fórmulas com o delta de Kronecker $\delta_{i,j}$ (lembrando que $m_{j,j} = 0$), temos, para todo $i \in S$,

$$m_{i,j} + \delta_{i,j} m_j = 1 + \sum_{k \in S} p_{i,k} m_{k,j}. \quad (*)$$

Multiplicando $(*)$ por π_i e somando em i , obtemos:

$$\sum_{i \in S} \pi_i m_{i,j} + \pi_j m_j = \sum_{i \in S} \pi_i + \sum_{i \in S} \sum_{k \in S} \pi_i p_{i,k} m_{k,j}.$$

Como $\sum_i \pi_i = 1$ e, pela invariância, $\sum_i \pi_i p_{i,k} = \pi_k$, a última soma fica

$$\sum_{k \in S} \pi_k m_{k,j}.$$

Assim,

$$\sum_{i \in S} \pi_i m_{i,j} + \pi_j m_j = 1 + \sum_{k \in S} \pi_k m_{k,j}.$$

Os termos $\sum_i \pi_i m_{i,j}$ e $\sum_k \pi_k m_{k,j}$ são idênticos e cancelam-se, restando

$$\pi_j m_j = 1.$$

Portanto,

$$m_j = \mathbb{E}_j[T_j^+] = \frac{1}{\pi_j},$$

que é exatamente a fórmula desejada. □

Corolário 7. *Se $(X_n)_{n \geq 0}$ é uma cadeia de Markov irredutível e aperiódica em um espaço de estados finito com matriz de transição $P = (p_{ij})$ então*

$$p_{ij}^n \xrightarrow{n \rightarrow \infty} \pi_j = \frac{1}{m_j}$$

para todo estado j . Ademais, o vetor $\pi = (\pi_1, \dots, \pi_n)$ é invariante.

Resumindo A cadeia “esquece” o ponto de partida; o comportamento a longo prazo é governado por π ; a fração do tempo em que o processo está em cada estado tende a π_i .

3 Teorema ergódico

O termo *ergodicidade* vem da mecânica estatística (Boltzmann, finais do século XIX). A *hipótese ergódica* afirmava, informalmente, que: *ao longo do tempo, um sistema físico isolado passa por todos os estados compatíveis com sua energia com igual frequência*. Matematicamente, isso significa que:

$$\text{Médias temporais} = \text{Médias espaciais (ou de ensemble)}$$

ou seja, para uma função f que mede alguma grandeza do sistema

$$\underbrace{\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k)}_{\text{média temporal observada}} = \underbrace{\sum_{i \in S} \pi_i f(i)}_{\text{média teórica estacionária}}.$$

Esse é exatamente o conteúdo do **Teorema Ergódico**: a média ao longo do tempo (na trajetória) coincide com a média segundo a distribuição estacionária (no espaço de estados). Um processo é ergódico se, observando-o por tempo suficientemente longo, obtém-se a mesma estatística que se veria olhando a distribuição estacionária. Em outras palavras:

- *Mistura*: perde a memória do início;
- *Estabiliza*: tem uma distribuição limite;
- *Representatividade*: uma trajetória longa contém informação sobre todo o sistema.

Este é o fundamento da mecânica estatística: pode-se substituir uma média no tempo (experimental) por uma média sobre o ensemble (teórica). Em linguagem de cadeias de Markov o *ensemble* corresponde à distribuição estacionária π ; a média temporal $\frac{1}{n} \sum_{k=0}^{n-1} f(X_k)$ converge quase certamente para $\mathbb{E}_\pi[f(X)] = \sum_x f(x) \pi(x)$, que é a média de ensemble.

Imagine um gás em uma caixa. Você pode:

- seguir uma única molécula por muito tempo (*média temporal*), ou
- tirar uma “foto” instantânea de milhares de moléculas (*média de ensemble*).

Se o sistema é ergódico, as duas descrições dão o mesmo resultado estatístico.

Teorema Ergódico para cadeias finitas, irredutíveis e aperiódicas

Seja $(X_n)_{n \geq 0}$ uma cadeia de Markov com matriz de transição $P = (p_{ij})$ em um espaço de estados finito S . Suponha que a cadeia é irredutível e aperiódica. Então:

(1) Existência e unicidade da distribuição estacionária Existe uma única distribuição de probabilidade $\pi = (\pi_i)_{i \in S}$ tal que

$$\pi P = \pi, \quad \sum_{i \in S} \pi_i = 1.$$

(2) Convergência em distribuição Para todo estado inicial $i \in S$, tem-se que a distribuição do estado X_n converge para a distribuição estacionária, independentemente do estado inicial:

$$\lim_{n \rightarrow \infty} p_{i,j}^n = \pi_j = \frac{1}{m_j}, \quad \forall j \in S$$

onde m_j é o tempo médio de retorno ao estado j .

(3) Lei dos Grandes Números ergódica Para qualquer função limitada $f : S \rightarrow \mathbb{R}$, o tempo médio gasto em cada estado converge quase certamente para a média com respeito à distribuição estacionária:

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow[n \rightarrow \infty]{q.c.} \sum_{i \in S} \pi_i f(i).$$

Demonstração da LGN. Escolha um estado fixo $r \in S$. Como a cadeia é irredutível em conjunto finito, o estado r tem tempo de retorno com esperança finita; seja

$$T_r^+ = \inf\{n \geq 1 : X_n = r\}, \quad m_r = \mathbb{E}[T_r^+ \mid X_0 = r] < \infty.$$

(Pela fórmula de Kac, $m_r = 1/\pi_r$.)

Defina as épocas de retorno sucessivas (épocas de regeneração)

$$\tau_0 = 0, \quad \tau_1 = T_r^+, \quad \tau_{k+1} = \inf\{n > \tau_k : X_n = r\} \quad (k \geq 1).$$

Para cada $k \geq 1$ considere o k -ésimo ciclo

$$\mathcal{C}_k = (X_{\tau_{k-1}}, X_{\tau_{k-1}+1}, \dots, X_{\tau_k-1})$$

e defina a duração e a soma de f nesse ciclo por

$$\xi_k := \tau_k - \tau_{k-1}, \quad S_k := \sum_{n=\tau_{k-1}}^{\tau_k-1} f(X_n).$$

Pela Propriedade Forte de Markov, os pares $(\xi_k, S_k)_k$ são i.i.d. quando a cadeia é iniciada em r . Como f é limitada, temos

$$\mathbb{E}[|S_1| \mid X_0 = r] \leq \|f\|_\infty \mathbb{E}[\xi_1 \mid X_0 = r] < \infty.$$

Para $n \geq 1$ seja

$$N(n) = \max\{k : \tau_k \leq n\}$$

o número de ciclos completos até tempo n . Então

$$\sum_{k=0}^{n-1} f(X_k) = \sum_{j=1}^{N(n)} S_j + R_n,$$

onde R_n é o resto (a soma de f sobre a porção incompleta entre $\tau_{N(n)}$ e $n - 1$). Temos a cota trivial

$$|R_n| \leq \|f\|_\infty (\tau_{N(n)+1} - \tau_{N(n)}) = \|f\|_\infty \xi_{N(n)+1}.$$

Aplicando a Lei Forte dos Grandes Números às sequências i.i.d. (S_j) e (ξ_j) , obtemos quase certamente:

$$\frac{1}{N(n)} \sum_{j=1}^{N(n)} S_j \xrightarrow[N(n) \rightarrow \infty]{} \mathbb{E}[S_1 \mid X_0 = r], \quad \frac{1}{N(n)} \sum_{j=1}^{N(n)} \xi_j = \frac{\tau_{N(n)}}{N(n)} \xrightarrow[N(n) \rightarrow \infty]{} \mathbb{E}[\xi_1 \mid X_0 = r] = m_r.$$

Dividindo as duas convergências (e lembrando que $N(n) \rightarrow \infty$ quase certamente quando $n \rightarrow \infty$) segue

$$\frac{\sum_{j=1}^{N(n)} S_j}{\tau_{N(n)}} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{\mathbb{E}[S_1 \mid X_0 = r]}{m_r}.$$

Como $\tau_{N(n)} \leq n < \tau_{N(n)+1}$, tem-se $\tau_{N(n)}/n \rightarrow 1$, e portanto

$$\frac{1}{n} \sum_{j=1}^{N(n)} S_j = \frac{\sum_{j=1}^{N(n)} S_j}{\tau_{N(n)}} \cdot \frac{\tau_{N(n)}}{n} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{\mathbb{E}[S_1 \mid X_0 = r]}{m_r}.$$

O resto satisfaz

$$\frac{|R_n|}{n} \leq \|f\|_\infty \frac{\xi_{N(n)+1}}{n} \leq \|f\|_\infty \frac{\xi_{N(n)+1}}{\tau_{N(n)}} \cdot \frac{\tau_{N(n)}}{n} \xrightarrow[n \rightarrow \infty]{\text{a.s.}} 0,$$

pois $\tau_{N(n)} \rightarrow \infty$ a.s. e $\xi_{N(n)+1}$ tem distribuição fixa com esperança finita. Assim

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow[n \rightarrow \infty]{\text{a.s.}} \frac{\mathbb{E}[S_1 \mid X_0 = r]}{m_r}.$$

Finalmente, identificamos a razão com a média estacionária: escrevendo

$$\mathbb{E}[S_1 \mid X_0 = r] = \sum_{i \in S} f(i) \mathbb{E} \left[\sum_{n=0}^{\tau_1-1} \mathbf{1}_{\{X_n=i\}} \mid X_0 = r \right],$$

a quantidade $\mathbb{E}[\sum_{n=0}^{\tau_1-1} \mathbf{1}_{\{X_n=i\}} \mid X_0 = r]$ é o número médio de visitas a i em um ciclo, e dividida pela duração média do ciclo m_r fornece a fração de tempo média gasta em i , que por Kac é exatamente π_i . Portanto

$$\frac{\mathbb{E}[S_1 \mid X_0 = r]}{m_r} = \sum_{i \in S} f(i) \pi_i,$$

concluindo a prova. □