Cadeias de Markov

Parte I. Introdução

Fernando Lucambio

Departamento de Estatística
Universidade Federal do Paraná

Julho, 2020

Estudaremos uma classe de processos aleatórios ou estocásticos que possuem uma determinada caracteríistica que pode ser descrita como perda de memória. Estes processos aparecem em inúmeras aplicações, incluindo sistemas de filas, redes de comunicação de computadores, sistemas biológicos e uma grande variedade de outras aplicações. Aplicações de Cadeias de Markov em medicina são bastante comuns e se tornaram uma ferramenta importante de tomada de decisão médica. Como resultado da sua ocorrência frequente, estes processos têm sido estudados extensivamente obtendo-se uma teoria rica a qual permite resolver os problemas relacionados com estes processos.

Um processo estocástico é um modelo matemático que evolui ao longo do tempo de forma probabilística e aqui vamos estudar um tipo especial de processo estocástico, chamado de Cadeia de Markov, onde o resultado de um experimento depende apenas do resultado do experimento anterior. Em outras palavras, o estado seguinte do sistema depende apenas do estado atual e não dos estados anteriores.

Definição (Processo aleatório)

Um processo aleatório é uma família $\{C_t\}_{t\in T}$ de variáveis aleatórias independentes ou não, definidas no mesmo espaço de probabilidade, indexadas pelo conjunto T.

Os processos estocástico podem ser classificados de diversas maneiras. Aqui estamos interessados em duas dessas classificações: a primeira é segundo as variáveis que o compõem, sejam discretas ou contínuas, identificando os processos estocásticos como discretos ou contínuos, respectivamente. Uma outra classificação é segundo o conjunto de índices. Caso o conjunto T seja um subconjunto dos números inteiros ou naturais, chamamos o processo estocástico de processo a tempo discreto, em outras situações chama-se de processo a tempo contínuo. Em qualquer caso, pensamos em um processo estocástico como uma família de variáveis aleatórias que evoluem com o tempo. Estamos preocupados com uma situação geral de modelos para a evolução aleatória. Um destes modelos satisfaz a seguinte propriedade: condicionado em seus valores no n-ésimo instante, seus valores futuros não dependem de seus valores anteriores. Esta propriedade provou ser muito útil na sua análise e é a teoria geral dos processos com essa propriedade à qual voltamos nossa atenção agora.

Considere um processo estocástico discreto $\{C_n\}$ com espaço amostral S, finito ou infinito enumerável. Pensemos $C_0, C_1, \cdots, C_{n1}$ como o passado, C_n como o presente e $C_{n+1}, C_{n+2}, \cdots$ como o futuro do processo em relação ao tempo n. A lei de evolução de um processo estocástico é frequentemente pensada em termos da distribuição condicional do futuro, dado o presente e os estados anteriores do processo.

Definição (Propriedade de Markov)

Seja $\{C_n\}$ um processo estocástico discreto com espaço amostral S, finito ou infinito enumerável. Dizemos que $\{C_n\}$ satisfaz a propriedade de Markov se dado o estado atual, os estados passados não têm influência sobre o futuro. A propriedade de Markov é definida precisamente pela exigência de que

$$P(C_{n+1} = c_{n+1} \,|\, C_0 = c_0, \cdots, C_n = c_n) = P(C_{n+1} = c_{n+1} \,|\, C_n = c_n),$$

para qualquer seja a escolha do número natural n e os números $c_0, c_1, \cdots, c_{n+1} \in S$. O espaço amostral S, de um processo estocástico discreto a tempo discreto, é chamado de espaço de estados.

Andrei Markov obteve os primeiros resultados para processos estocásticos discretos finitos em 1906. Uma generalização para espaços de estados infinitos enumeráveis foi dada por Kolmogorov em 1936. Andrei Nikolaevich Kolmogorov (1903-1987) foi um matemático soviético. Kolmogorov participou das principais descobertas científicas do século XX nas áreas de probabilidade e estatística e na teoria da informação. A definição desta propriedade, também chamada de memória Markoviana, é que os estados anteriores são irrelevantes para a predição dos estados seguintes, desde que o estado atual seja conhecido.

Definição

Um processo estocástico $\{C_n\}$ discreto satisfaz a propriedade de Markov se, para cada n e m, a distribuição condicional de $C_{n+1}, \cdots, C_{n+m}$ dado $C_0, C_1, \cdots, C_n$ é a mesma que a sua distribuição condicional dado C_n.

Isto quer dizer que, um processo estocástico satisfazendo a propriedade de Markov satisfaz que

$$P(C_{n+1}, \cdots, C_{n+m} \,|\, C_0, C_1, \cdots, C_n) = P(C_{n+1}, \cdots, C_{n+m} \,|\, C_n)\cdot$$

Definição (Cadeias de Markov)

Um processo estocástico $\{C_n\}$ que satisfaz a propriedade de Markov é chamado de processo de Markov. Se, além disso, o processo estocástico for a tempo discreto e formado por variáveis aleatórias discretas o processo de Markov é chamado Cadeia de Markov.

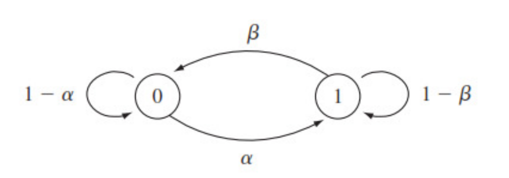
Observemos que ao definirmos Cadeia de Markov nada é dito acerca do espaço de estados. Agora, como as variáveis aleatórias $C_0, C_1, \cdots, C_n, \cdots$ que formam a Cadeia de Markov são discretas, então o espaço de estados S é finito ou infinito enumerável.

O estudo das Cadeias de Markov é válido a partir de dois pontos de vista. Em primeiro lugar, existe uma ampla teoria desenvolvida e, em segundo lugar, há um grande número de sistemas que surgem na prática que podem ser modelados desta forma, de modo existirem muitas aplicações.

Como suporte computacional utilizamos a linguagem de programação e ambiente de desenvolvimento integrado para cálculos estatísticos e gráficos R, última versão 3.5.2, Eggshell Igloo de 20 de dezembro de 2018.

Exemplo (Cadeias de Markov)

Uma Cadeia de Markov com dois estados é um processo de Markov para um sistema que pode assumir somente dois valores, por exemplo, 0 e 1. Podemos observar um gráfico representativo abaixo (grafo). Partindo do estado 0, permanece nele com probabilidade $1-\alpha$ e assume valor 1 com probabilidade α. Da mesma forma, se o estado atual é 1, permanece nele com probabilidade $1-\beta$ e muda para 0 com probabilidade β.



Grafo das probabilidades de transição na cadeia com dois estados.

Definição (Probabilidade de transição)

Seja $\{C_t \ : \ t \in \mathbf{N}\}$ uma Cadeia de Markov com espaço de estados S e sejam $x, y \in S$. A probabilidade

$$P(C_{t+1} = y \,|\, C_t = x),$$

se conhece como probabilidade de transição em um passo ou simplesmente probabilidade de transição. Também denotada como $\gamma_{x,y}(t, t+1)$, a qual representa a probabilidade de transição do estado x no tempo t ao estado y no tempo $t+1$.

Podemos encontrar processos nos quais as probabilidades de transição variam com o tempo e, portanto, necessitam ser explicitamente escritas como uma função do tempo t, mas não consideraremos tais processos no presente texto e doravante, presume-se que a probabilidade de transição é independente do tempo.

Definição (Cadeia de Markov homogênea)

Uma Cadeia de Markov $\{C_t \ : \ t \in \mathbf{N}\}$ é dita ser homogênea ou estacionária se as probabilidades de transição $\gamma_{x,y}(t)$ não dependem do tempo.

Estudamos agora três características importantes das Cadeias de Markov: a função de transição, a distribuição inicial e a matriz de transição. Toda vez que lidemos com situações que possam ser modeladas desta maneira, estaremos interessados em identificar estas características. Mais ainda, estas características serão importantes para encontrar propriedades das Cadeias de Markov.

Função de transição

Pela definição de função de transição podemos perceber que a probabilidade de transição, numa Cadeia de Markov estacionária, é uma função dos estados e não mais dos instantes de tempo. Dedicaremos especial atenção a esta função, a qual permitirá deduzir propriedades destes modelos.

Definição (Função de transição).

Seja $\{C_t \ : \ t \in \mathbf{N}\}$ uma Cadeia de Markov estacionária com espaço de estados S. A função γ, definida como

$$\gamma_{x,y} \ = P(C_1 = y \,|\, C_0 = x), \qquad x, y \in S,$$

é chamada de função de transição da cadeia.

Em particular, dado que a cadeia satisfaz as exigências da definição de cadeia estacionária, podemos escrever que

$$P(C_{t+1} = y \,|\, C_t = x) \;=\; \gamma_{x,y}\,, \qquad t \geq 1\cdot$$

Agora, pela propriedade Markoviana, temos que

$$P(C_{t+1} = y \,|\, C_0 = c_0, \cdots, C_{t1} = c_{t1}, C_t = x) \;=\; \gamma_{x,y}\cdot$$

Em outras palavras, se a Cadeia de Markov está no estado x no tempo n, então não importa como ela chegou a x, ela tem probabilidade $\gamma_{x,y}$ de estar no estado y no passo seguinte. Por esta razão os números $\gamma_{x,y}$ são chamados também de probabilidades de transição de uma etapa da Cadeia de Markov. Esta função, a qual é uma probabilidade condicional, satisfaz propriedades básicas:

(a) $\gamma_{x,y} \geq 0, x, y \in S.$

(b) $\displaystyle\sum_{y \in S} \gamma_{x,y} = 1, x \in S.$

Distribuição inicial

Um vetor que consiste de números não negativos que somam 1 é chamado de vetor de probabilidades. Um vetor de probabilidades cujas coordenadas especificam as probabilidades de que uma Cadeia de Markov esteja em cada um dos seus estados no tempo inicial é chamado de distribuição inicial da cadeia ou o vetor de probabilidade inicial.

Definição (Distribuição inicial).

Seja $\{C_t \, : \, t \in \mathbf{N}\}$ uma Cadeia de Markov estacionária com espaço de estados S. A função $\pi_0(x)$, $x \in S$, definida como

$$\pi_0(x) \, = \, P(C_0 = x), \qquad x \in S,$$

é chamada de distribuição inicial da cadeia.

Fica implícito que a dimensão do vetor π_0 é igual ao número de estados ou elementos em S.

Matriz de transição

Suponha uma Cadeia de Markov estacionária com espaço de estados

$$S = \{s_1, s_2, s_3, \cdots, s_d\},$$

finito. Em situações como esta é conveniente definir a probabilidade de transição como a matriz

$$\Gamma = \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ \vdots \\ s_d \end{array} \begin{array}{c} \begin{array}{ccccc} s_1 & s_2 & s_3 & \cdots & s_d \end{array} \\ \begin{pmatrix} \gamma_{s_1,s_1} & \gamma_{s_1,s_2} & \gamma_{s_1,s_3} & \cdots & \gamma_{s_1,s_d} \\ \gamma_{s_2,s_1} & \gamma_{s_2,s_2} & \gamma_{s_2,s_3} & \cdots & \gamma_{s_2,s_d} \\ \gamma_{s_3,s_1} & \gamma_{s_3,s_2} & \gamma_{s_3,s_3} & \cdots & \gamma_{s_3,s_d} \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_{s_d,s_1} & \gamma_{s_d,s_2} & \gamma_{s_d,s_3} & \cdots & \gamma_{s_d,s_d} \end{pmatrix} \end{array}$$

de elementos as probabilidades de transição entre os estados, escritas como

$$\gamma_{x,y} = P(X_1 = y | X_0 = x), \qquad x, y \in S\cdot$$

Definição (Matriz de transição).

Seja $\{C_n\}$, uma Cadeia de Markov estacionária com $S = \{s_1, s_2, s_3, \cdots, s_d\}$ o espaço de estados finito. Então a matriz quadrada Γ, dada acima, é conhecida como a matriz de probabilidades de transição.

Uma Cadeia de Markov $\{C_n\}$ estacionária com espaço de estados S. A matriz $\Gamma = (\gamma_{x,y})$, de probabilidades de transição, satisfaz as seguintes propriedades:

(a) $\gamma_{x,y} \geq 0$,

(b) $\displaystyle\sum_{y \in S} \gamma_{x,y} = 1$.

Este último resultado significa que a partir de qualquer estado x, com probabilidade finita, uma cadeia assume necessariamente algum elemento do espaço de estado na próxima vez. Geralmente uma matriz quadrada satisfazendo estas duas propriedades se disse que é uma matriz estocástica. Devido à propriedade de Markov, esta matriz capta a essência do processo, e determina o comportamento da cadeia em qualquer momento no futuro.

Exemplo (Motor de busca do Google)

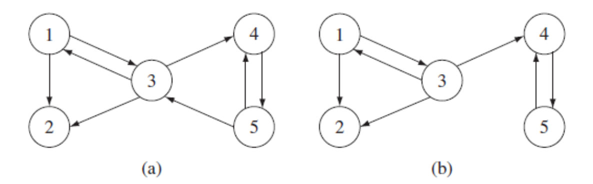
Os motores de busca surgiram logo após o aparecimento da Internet, com a intenção de prestar um serviço extremamente importante: a busca de qualquer informação na rede, apresentando os resultados de uma forma organizada e também com a proposta de fazer isto de uma maneira rápida e eficiente. A partir deste preceito básico, diversas empresas se desenvolveram, chegando algumas a valer bilhões de dólares. Entre as maiores empresas encontram-se Google, Yahoo, Lycos e outras. Os buscadores se mostraram imprescindíveis para o fluxo de acesso e a conquista de novos visitantes.

Matriz de probabilidades do grafo apresentado abaixo:

$$\Gamma = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \!\!\begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \\ 1/3 & 1/3 & 0 & 1/3 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix} \end{array} \cdot$$

Exemplo (Motor de busca do Google, continuação)

Suponha que o internauta navega por páginas da Web em um universo de cinco páginas, como mostrado na figura abaixo (a), sendo cada página os elementos do espaço de estados $S = \{1, 2, 3, 4, 5\}$. O internauta escolhe a próxima página para ver selecionando com igual probabilidade a partir das páginas apontadas pela página atual. Se uma página não tem qualquer ligação de saída (por exemplo, página 2), em seguida, o interessado seleciona qualquer uma das páginas do universo, com igual probabilidade. Poderíamos estar interessados em encontrar a probabilidade de que o internauta veja a *i*-ésima página.

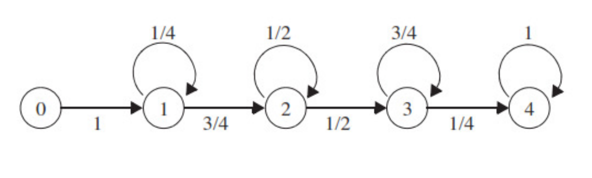


Grafo das probabilidades de transição num buscador.

Exemplo (Fast food)

Suponha que cada vez que uma criança adquire uma refeição de criança em seu restaurante fast food favorito, ele recebe uma das quatro figuras de super-heróis. Naturalmente, a criança quer coletar todas as quatro figuras de ação e assim ele come regularmente no restaurante para completar a coleção. Este processo pode ser descrito por uma Cadeia de Markov e a matriz de probabilidades de transição é da forma:

$$\Gamma = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1/4 & 3/4 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 \\ 0 & 0 & 0 & 3/4 & 1/4 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}.$$



Grafo das probabilidades de transição do Exemplo Fast food.

Exemplo (Fast food, continuação)

Neste caso, seja $S = \{0, 1, 2, 3, 4\}$ o espaço de estados, ou seja, o número das diferentes figuras de super-heróis que a criança tem recolhido após a compra de k refeições. Supondo que cada refeição contém um dos quatro super-heróis, com igual probabilidade, e que a matriz de transição em qualquer refeição é independente do que está contido em todas as refeições anteriores ou futuras, então a matriz de probabilidade de transição é mostrada anteriormente.

Inicialmente, quer dizer, antes de todas as refeições serem compradas, o processo começa em 0, estado no qual a criança não tem figuras de super-heróis. Quando a primeira refeição é comprada, a Cadeia de Markov deve passar para o estado um, uma vez, não importa qual figura de ação está contido na refeição, a criança terá agora um super-herói. Assim, $\gamma_{0,1} = 1$, e $\gamma_{0,j} = 0$ para todo $j \neq 1$. Se a criança tem uma figura de ação, quando ele compra a próxima refeição, ele tem uma chance de 25% de receber um duplicado e 75% de chance de conseguir uma nova figura de super-herói.

Assim, $\gamma_{1,1} = 1/4$, $\gamma_{1,2} = 3/4$ e $\gamma_{1,j} = 0$ para a $j \neq 1, 2$. Lógica semelhante é usada para completar o restante da matriz. A criança pode estar interessada em saber o número médio de refeições que ela precisa comprar até que sua coleção esteja completa. Ou, talvez a criança salvou-se apenas o dinheiro suficiente para comprar 10 almoços e quer saber quais suas chances de completar o conjunto antes de ficar sem dinheiro. Vamos desenvolver a teoria necessária para responder a essas perguntas.