ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

~~~~~~*~~~~~~

# BÁO CÁO ĐỒ ÁN THỰC HÀNH CUỐI KÌ: ĐỀ 3 - XÍCH MARKOV

**Sinh viên thực hiện:** Đàm Tử Tâm

**Mã số sinh viên:** 21120551

**Môn học:** Toán ứng dụng & thống kê

**Lớp:** 21_2

**Thành phố Hồ Chí Minh – 2023**
~~~~~~

# I. Thông tin cá nhân:

Sinh viên thực hiện: Đàm Tử Tâm

Mã số sinh viên: 21120551

Lớp: 21_2

# II. Nội dung thực hiện:

**Đề 3 - xích Markov:** ứng dụng xích Markov vào 1 bài toán cụ thể trong thực tế qua việc dự báo thời tiết ngày hôm sau dựa trên thời tiết hiện tại.

# III. Lý thuyết xích Markov:

## 1. Ví dụ dẫn nhập:

Giả sử một hệ thống vật lý hoặc toán học trải qua quá trình thay đổi sao cho tại bất kỳ thời điểm nào, nó có thể có một trong số các trạng thái hữu hạn. Ví dụ, thời tiết trong một thành phố có thể ở một trong ba trạng thái: nắng, sương mù, hoặc mưa. Hoặc một người có thể ở một trong bốn trạng thái cảm xúc: vui vẻ, buồn, tức giận hoặc lo lắng. Giả sử rằng hệ thống thay đổi theo thời gian từ trạng thái này sang trạng thái khác và vào các thời điểm đã được định sẵn thì trạng thái của hệ thống sẽ được quan sát. Nếu trạng thái của hệ thống tại một thời điểm quan sát nào đó không thể được dự đoán chính xác, nhưng xác suất một trạng thái cụ thể của hệ thống sẽ xảy ra có thể được dự đoán khi đã biết trạng thái của hệ thống tại lần quan sát trước đó, thì quá trình thay đổi này được gọi là một xích Markov.

## 2. Định nghĩa 1:

- Định nghĩa: nếu một xích Markov có k trạng thái có thể xảy ra, tạm gọi là 1, 2, … , k thì xác suất hệ thống ở trạng thái i tại bất kỳ thời điểm quan sát nào sau khi nó ở trạng thái j tại thời điểm quan sát trước đó sẽ được ký hiệu là $p_{ij}$ và được gọi là **xác suất chuyển trạng thái** từ trạng thái j sang trạng thái i. Ma trận $P = [p_{ij}]$ được gọi là **ma trận chuyển trạng thái** của xích Markov.

- Ví dụ, trong một xích Markov có 3 trạng thái (1, 2, 3) thì ma trận chuyển trạng thái của xích như sau:

***Trạng thái hiện tại***

$$\begin{array}{cccc} \textbf{\textit{1}} & \textbf{\textit{2}} & \textbf{\textit{3}} & \\ \multicolumn{3}{c}{\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}} & \begin{matrix} \textbf{\textit{1}} \\ \textbf{\textit{2}} \\ \textbf{\textit{3}} \end{matrix} \end{array} \quad \textbf{\textit{Trạng thái mới}}$$

- Trong ma trận trên, $p_{32}$ là xác suất mà hệ thống sẽ chuyển từ trạng thái 3 sang trạng thái 2, $p_{11}$ là xác suất mà hệ thống sẽ ở trong trạng thái 1 nếu trước đó hệ thống vẫn ở trạng thái 1, tương tự với các phần tử còn lại.

## 3. Định nghĩa 2:

- Định nghĩa: **vector trạng thái** của một xích Markov có k trạng thái là một vector cột x gồm k phần tử, với phần tử thứ i của vector được ký hiệu là $x_i$ là xác suất hệ thống ở trạng thái thứ i tại thời điểm quan sát đó.
- Từ định nghĩa có thể rút ra được các phần tử trong bất kỳ vector trạng thái nào của một xích Markov đều không âm và tổng các phần tử của một vector trạng thái bằng 1 (do là xác suất).

### Định lý 1:

- Nếu P là ma trận chuyển trạng thái của một xích Markov và $x^{(n)}$ là một vector trạng thái của quan sát thứ n thì $\mathbf{x^{(n+1)} = Px^{(n)}}$, chứng minh của định lý này liên quan đến các ý tưởng từ lý thuyết xác suất nên sẽ không được trình bày ở đây. Từ định lý này, ta có thể suy ra được:

  $x^{(1)} = Px^{(0)}$

  $x^{(2)} = Px^{(1)} = P^2x^{(0)}$

  $x^{(3)} = Px^{(2)} = P^3x^{(0)}$

  $\vdots$

  $x^{(n)} = Px^{(n-1)} = P^nx^{(0)}$

  Bằng cách này, ta cần vector trạng thái ban đầu $x^{(0)}$ và ma trận chuyển trạng thái P để xác định $x^{(n)}$, với n = 1, 2, ….

## 4. Định nghĩa 3:

- Định nghĩa: cho P là một ma trận chuyển trạng thái, P được gọi là chính quy nếu tồn tại $n \in \mathbb{N}$ sao cho mọi số hạng của $P^n$ đều là số dương.
- Một xích Markov có ma trận chuyển trạng thái P là chính quy gọi là **xích Markov chính quy.** Mỗi xích Markov chính quy có một vector trạng thái q cố định sao cho $P^n x^{(0)}$ sẽ tiến gần đến q khi n càng tăng, với vector trạng thái $x^{(0)}$ bất kỳ. Kết quả này có ý nghĩa quan trọng trong lý thuyết xích Markov và dựa trên định lý sau đây:

### Định lý 2:

Cho P là một ma trận chuyển trạng thái, nếu P chính quy thì với $n \to \infty$,

$$P^n \to \begin{bmatrix} q_1 & q_1 & \cdots & q_1 \\ q_2 & q_2 & \cdots & q_2 \\ \vdots & \vdots & & \vdots \\ q_k & q_k & \cdots & q_k \end{bmatrix}$$

trong đó, $q_i$ là một số dương và $q_1 + q_2 + \ldots + q_k = 1$

### Định lý 3:

Cho P là một ma trận chuyển trạng thái và x là một vector trạng thái bất kỳ, nếu P chính quy thì với $n \to \infty$,

$$P^n x \to \begin{bmatrix} q_1 \\ q_2 \\ \vdots \\ q_k \end{bmatrix} = q$$

với q được gọi là **vector phân phối dừng**, q không phụ thuộc vào n và tất cả các phần tử của q đều là số dương

### Định lý 4:

Vector phân phối dừng q của ma trận chuyển trạng thái chính quy P là một vector xác suất duy nhất thỏa mãn phương trình **Pq = q**.

## 5. Xác suất xích Markov:

Gọi $X_t$ là biến ngẫu nhiên chỉ trạng thái của hệ thống tại thời điểm t, ta có :

- $P(X_{t+k} = i \mid X_{t+k-1} = j, X_{t+k-2} = m, \ldots) = P(X_{t+k} = i \mid X_{t+k-1} = j)$

- $P(X_{t+k} = i \mid X_t = j) = P(X_k = i \mid X_0 = j)$

# IV. Ứng dụng của xích Markov vào việc dự báo thời tiết ngày hôm sau dựa trên thời tiết hiện tại:

Xích Markov được ứng dụng trong nhiều lĩnh vực giúp cải thiện đời sống như phân tích ngôn ngữ (dự đoán chuỗi từ), kinh tế (dự đoán xu hướng thị trường, quản lý rủi ro tài chính),…. Trong đó, việc dự báo thời tiết hằng ngày cũng được áp dụng xích Markov để vận hành.

Lấy dữ liệu về tỉ lệ nắng/mưa trong một ngày của tháng một ở thành phố Cape Coast – nước Ghana được nghiên cứu bởi Meshach Tettey, Francis T. Oduro, David Adedia và Daniel A. Abaye trong bài báo "Markov chain analysis of the rainfall patterns of five geographical locations in the south eastern coast of Ghana" như sau: nếu thời tiết ngày hiện tại của thành phố là mưa, tỉ lệ ngày mai trời vẫn mưa là 6%, tỉ lệ ngày mai trời nắng là 94% ; nếu thời tiết ngày hôm nay của thành phố là nắng, tỉ lệ ngày mai trời mưa là 16%, tỉ lệ ngày mai trời vẫn nắng là 84%. Dựa vào thông tin đó, ta lập được ma trận chuyển trạng thái như sau:

***thời tiết ngày hôm nay***

$$P = \begin{array}{c} \textit{mưa} \quad \textit{nắng} \\ \begin{bmatrix} 0.06 & 0.16 \\ 0.94 & 0.84 \end{bmatrix} \end{array} \begin{array}{l} \textit{mưa} \\ \textit{nắng} \end{array} \quad \textit{thời tiết ngày hôm sau}$$

Giả sử hôm nay thành phố mưa, ta có vector trạng thái ban đầu như sau:

$$x^{(0)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{array}{l} \textit{mưa} \\ \textit{nắng} \end{array}$$

Để dự báo thời tiết ngày mai (sau 1 ngày kể từ ngày hiện tại), ta dùng định lý 1:

$$x^{(1)} = Px^{(0)} = \begin{bmatrix} 0.06 & 0.16 \\ 0.94 & 0.84 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.06 \\ 0.94 \end{bmatrix}$$

Dựa vào vector trạng thái $x^{(1)}$, có thể thấy được nếu thời tiết ngày hiện tại của thành phố là mưa, tỉ lệ mưa ngày mai là 6%, tỉ lệ nắng là 94%. Ta tiếp tục dùng định lý 1 để dự báo thời tiết sau 2 ngày kể từ ngày hiện tại:

$$x^{(2)} = Px^{(1)} = \begin{bmatrix} 0.06 & 0.16 \\ 0.94 & 0.84 \end{bmatrix} \begin{bmatrix} 0.06 \\ 0.94 \end{bmatrix} = \begin{bmatrix} 0.154 \\ 0.846 \end{bmatrix}, \text{ hoặc}$$

$$x^{(2)} = P^2 x^{(0)} = \begin{bmatrix} 0.154 & 0.144 \\ 0.846 & 0.856 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.154 \\ 0.846 \end{bmatrix}$$

Dựa vào vector trạng thái $x^{(2)}$, có thể thấy được nếu thời tiết ngày hiện tại của thành phố là mưa thì sau 2 ngày kể từ ngày hiện tại, tỉ lệ mưa là 15.4%, tỉ lệ nắng là 84.6%. Áp dụng tương tự đối với $x^{(3)}, x^{(4)}, \ldots$ để dự báo thời tiết sau 3, 4,… ngày kể từ ngày hiện tại ở thành phố Cape Coast.

Ngoài ra, ta thấy P là chính quy do $P^2 = \begin{bmatrix} 0.154 & 0.144 \\ 0.846 & 0.856 \end{bmatrix}$ có các phần tử đều là số dương, điều này đồng nghĩa với việc xích Markov sẽ có vector phân phối dừng q. Áp dụng định lý 4, ta tìm q như sau:

$$Pq = q$$
$$\Leftrightarrow Pq - q = 0$$
$$\Leftrightarrow (P - I_2)q = 0 \ (*)$$

đặt q = $\begin{pmatrix} a \\ b \end{pmatrix}$, A là ma trận mở rộng sau khi ma trận hóa hệ phương trình $(*)$, ta có:

$$\text{A} = \left(\begin{array}{cc|c} -0.94 & 0.16 & 0 \\ 0.94 & -0.16 & 0 \end{array}\right) \to \left(\begin{array}{cc|c} 1 & -8/47 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

Hệ phương trình có vô số nghiệm, với $\begin{cases} a = \frac{8\alpha}{47} \\ b = \alpha \end{cases}$, vậy q = $\begin{pmatrix} \frac{8\alpha}{47} \\ \alpha \end{pmatrix}$

Từ kết quả rút ra được ở định nghĩa 2, ta có $\frac{8\alpha}{47} + \alpha = 1 \Leftrightarrow \alpha = \frac{47}{55}$, thế $\alpha$ vào q, ta được q = $\begin{pmatrix} 0.145 \\ 0.855 \end{pmatrix}$. *Như vậy có thể* kết luận: về lâu dài, tỉ lệ mưa trong một ngày của tháng một ở thành phố Cape Coast là 14.5%, tỉ lệ nắng là 85.5%.

Phía trên là một ví dụ thực tế được nghiên cứu bởi Meshach Tettey, Francis T. Oduro, David Adedia và Daniel A. Abaye cho thấy việc xích Markov đã được ứng dụng vào lĩnh vực dự báo thời tiết trong đời sống. Tất cả các dữ liệu đều là các dữ liệu thực tế trong khoảng thời gian 1980 – 2010 được các tác giả thu thập và xử lý. Ngoài Cape Coast, bài báo còn trình bày thêm về nghiên cứu tỉ lệ nắng/mưa trong một ngày của từng tháng tại các khu vực khác thuộc nước Ghana là Accra, Akuse, Akatsi và Keta bằng xích Markov, từ đó suy ra được tỉ lệ nắng mưa trong từng tháng (áp dụng phân phối dừng) và xác định mùa mưa và mùa nắng trong từng khu vực. Việc này giúp ích cho các ngành nông nghiệp, những người hành nghề bảo hiểm nông nghiệp và các tổ chức kinh doanh khác trong khu vực, giúp họ có thể lập kế hoạch hoạt động phù hợp; chuẩn bị đất và giống cây trồng, đồng thời đưa ra lời khuyên cho khách hàng của họ về các khoản đầu tư. Nghiên cứu của họ cũng là nghiên cứu đầu tiên ở Ghana chứng minh ứng dụng thực tế của xích Markov đối với dữ liệu khí tượng.

# V. Minh họa xích Markov bằng python:

Chương trình minh họa xích Markov bằng python gồm các hàm và dữ liệu sau:

- Hàm deepcopy: dùng để sao chép ma trận input ra một ma trận mới và trả về ma trận này trong output.
- Hàm print_matrix: input là một ma trận cùng một list các trạng thái của xích Markov, in ra màn hình ma trận input cùng với tiêu đề các dòng/các cột của ma trận input dựa trên list trạng thái input.
- Hàm multi_matrix: nhân hai ma trận input (A x B) và trả về ma trận kết quả trong output. Đầu tiên kiểm tra nếu số cột của ma trận A khác số dòng của ma trận B thì in ra màn hình không thể nhân hai ma trận và kết thúc hàm. Ngược lại, dùng 3 vòng lặp for lồng nhau để nhân hai ma trận: xét từ vòng for ngoài cùng vào, vòng for i lần lượt duyệt qua từng dòng của A, rồi vòng for j lồng phía trong duyệt từng cột của B và vòng for trong cùng là vòng for k duyệt từng cột của A. Trong vòng for k, lấy tổng A[i][k] x B[k][j]. Sau khi chạy hết vòng for k, ta được phần tử C[i][j] trong ma trận kết quả.
- Hàm markov_chain: input là một ma trận chuyển trạng thái A, vector cột trạng thái ban đầu v và số ngày sau ngày hiện tại cần dự báo thời tiết. Hàm sẽ trả về vector kết quả tính toán $A^{day}v$ để xác định xác suất xảy ra các trạng thái thời tiết sau day ngày kể từ ngày hiện tại. Nếu day có giá trị 0, hàm trả về vector v là vector trạng thái ban đầu và ngưng hàm. Ngược lại, tính $A^{day}v$ bằng hàm multi_matrix rồi trả về vector kết quả.
- Hàm check_positive_matrix: input là ma trận A, hàm sẽ kiểm tra ma trận có chứa toàn số dương hay không, nếu có bất kì phần tử nào không dương trong ma trận thì trả về False, ngược lại trả về True.
- Hàm check_regular_matrix: input là ma trận A, hàm kiểm tra ma trận có là chính quy hay không bằng cách lần lượt xét từ $A^2$ đến $A^{52}$ (dùng hàm multi_matrix) xem có bất kì trường hợp nào là ma trận chứa toàn số dương hay không. Nếu không có, hàm trả về False và coi như ma trận không chính quy, ngược lại trả về True và ma trận là chính quy.
- Hàm steady_state_vec: input là ma trận chuyển trạng thái A1, hàm sẽ trả về vector phân phối dừng V. Đầu tiên kiểm tra A1 có là ma trận chính quy hay không (bằng hàm check_regular_matrix), nếu không thì ngưng hàm và trả về giá trị None, ngược lại thì

tiếp tục. Đầu tiên, dùng hàm deepcopy_matrix để copy ma trận A1 và gán vào A (nhằm tránh thay đổi giá trị của ma trận input). Sau đó, thực hiện giải thuật Gauss với ma trận A:

  - Với cột đang xét:
    - Nếu chứa toàn bộ số 0 → sang cột kế tiếp
    - Chứa ít nhất một số hạng khác 0 → sang bước 2. Việc kiểm tra trên được thực đánh dấu bằng biến flag trong hàm (nếu flag True thì sang bước 2, False thì sang cột kế tiếp)
  - Nếu phần tử đầu cột = 0 → hoán vị dòng này với dòng bên dưới có phần tử trong cột khác 0
  - Nhân dòng 1 cho 1/a (a là số hạng đầu dòng)
  - Cộng các dòng còn lại (khác dòng đang ở phần tử đầu cột) cho một số thích hợp cho dòng đầu để biến đổi = 0
  - Lặp lại các bước trên với ma trận đã che dòng đã làm

Lúc này, ma trận A sẽ có dòng cuối toàn là số 0, nếu ghép vector cột B (toàn 0) vào A thì A sẽ trở thành hệ phương trình Ax=0 có vô số nghiệm với 1 biến tự do. Lúc này, không thực hiện ghép B vào A mà sẽ tìm các hệ số đứng trước biến tự do của các nghiệm của hệ phương trình bằng cách đưa các phần tử của cột cuối vào vector V trừ dòng cuối của cột cuối, sau đó đưa 1 vào V (hệ số 1 trước biến tự do). Mà ta có tổng các phần tử của vector này bằng 1 (do là vector xác suất), lúc này tính giá trị của biến tự do bằng cách lấy 1/(tổng tất cả các phần tử của vector) và gán vào sum, sau đó nhân sum cho từng phần tử của vector V, ta được vector phân phối dừng V và hàm sẽ trả về vector này.

- Hàm print_ans: hàm in ra màn hình kết quả của từng trường hợp test, bao gồm in và giải thích ma trận chuyển trạng thái, in ra thời tiết hôm nay, in ra dự báo thời tiết của những ngày hôm sau, in ra dự báo thời tiết về lâu dài (phân phối dừng).
- Dữ liệu test: là các dữ liệu thực tế được lấy tài liệu tham khảo "Markov chain analysis of the rainfall patterns of five geographical locations in the south eastern coast of Ghana" của các tác giả Meshach Tettey, Francis T. Oduro, David Adedia & Daniel A. Abaye nghiên cứu về tỉ lệ mưa/nắng ở 5 khu vực của nước Ghana: Cape Coast, Accra, Akuse, Akatsi và Keta. Các dữ liệu được thu thập từ năm 1980 đến 2010.

# VI. Tài liệu tham khảo:

- Tài liệu Toán ứng dụng & thống kê - chương xích Markov - Đại học Khoa học Tự Nhiên, ĐHQG Thành phố Hồ Chí Minh
- “Markov chain analysis of the rainfall patterns of five geographical locations in the south eastern coast of Ghana” - Meshach Tettey, Francis T. Oduro, David Adedia & Daniel A. Abaye https://earth-perspectives.springeropen.com/articles/10.1186/s40322-017-0042-6
- Elementary Linear Algebra: Applications Version, 11th Edition - Howard Anton, Chris Rorres https://industri.fatek.unpatti.ac.id/wp-content/uploads/2019/03/037-Elementary-Linear-Algebra-Applications-Version-Howard-Anton-Chris-Rorres-Edisi-1-2013.pdf