Xác Suất Thông Kê

Phạm Tấn Anh Quân

Đại Học Khoa Học Tự Nhiên, Đại Học Quốc Gia Tp.HCM

Subject Index Probability, statistics.

1. Giới thiệu

Trong thế giới vận động không ngừng thì không có gì là hoàn toàn chắc chắn: Liệu máy bay có bị rơi? Liệu số dân có vừa đủ với lượng việc làm? Liệu một dự án có thành công? Đó là những câu hỏi mà không chỉ các nhà Toán học mà hầu như ai trong tất cả chúng ta đều phải đối mặt hằng ngày. Để hỗ trợ cho những câu hỏi này, các nhà Toán học đã phát minh ra một công cụ đó là **Xác suất Thống Kê**.

Xác suất có thể hiểu là khả năng một vật, hay một sự kiện có xuất hiện/ diễn ra hay không. *Thống kê* là cách con người ghi chép lại những số liệu, tần suất xuất hiện của một vật, diễn biến của một hoặc nhiều cá thể trong cùng một môi trường, ...

Chúng ta dựa vào những con số thống kê để tìm ra một quy luật nhằm tính xác suất một sự vật hay một sự kiện sẽ diễn ra, đó là **Xác suất Thống kê**. Tuy không giúp chúng ta biết chắc chắn một sự vật có diễn ra hay không, nhưng nó giúp chúng ta hệ thống hóa được toàn bộ các diễn biến có thể của một sự vật/ sự kiện, và từ đó đưa ra một mô hình/ bài toán để giải quyết những vấn đề sau đó.

2. Khái niệm cơ bản

2.1. *Phép thử và sự kiện (Experiment and Event)*

Phép thử được hiểu là thực hiện một tập hợp điều kiện xác định, có thể là quan sát hoặc thí nghiệm trên một hiện tượng nào đó. *Sự kiện sơ cấp*, hay *biến cố sơ cấp* là kết quả thu được sau một lần sử dụng phép thử. Chúng ta có thể gọi tắt là *kết quả*. *Sự kiện* là tập hợp của một hoặc nhiều các sự kiện sơ cấp dựa trên một tiêu chí nào đó. Tập hợp tất cả các kết quả của phép thử, hay tất cả các sự kiện sơ cấp, chúng ta gọi là *không gian sự kiện sơ cấp* (hoặc *không gian mẫu*) của phép thử đó, và kí hiệu là Ω. Mỗi kết quả ω của phép thử là kết quả thuận lợi đối với *sự kiện* A nếu như A xảy ra khi kết quả của phép thử đó là ω.

Ví dụ 2.1. Gieo một con xúc xắc xem mặt trên cùng ra số nào, đây là một **phép thử C**. Các kết quả $k = 1, 2, 3, 4, 5, 6$, với k là số có thể xuất hiện trên mặt trên cùng của xúc xắc, là các **sự kiện sơ cấp** của phép thử C. **Không gian sự kiện sơ cấp** của C là $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Ta có thể xét một số **sự kiện** đối với phép thử C như sau: *Sự kiện A là sự kiện mặt trên cùng là số bé hơn hoặc bằng 2. Sự kiện B là sự kiện mặt trên cùng là số chẵn.*

Các kết quả thuận lợi của A là $\{1, 2\}$. Các kết quả thuận lợi của B là $\{2, 4, 6\}$.

Như vậy, mỗi sự kiện A có thể được coi như là một tập con của không gian sự kiện sơ cấp Ω bao gồm các kết quả thuận lợi đối với A. Khi nói đến sự kiện A ta coi nó như là một tập hợp hoặc là tính chất của sự kiện đó. Kết quả của một phép thử thỏa mãn sự kiện đồng nghĩa với việc một phần tử thuộc một tập hợp nào đó.

Có hai loại *sự kiện* đặc biệt sau

- *Sự kiện chắc chắn* là sự kiện luôn xảy ra với mọi phép thử. Sự kiện này, nếu được coi như là một tập con, trùng với không gian mẫu Ω.
- *Sự kiện không thể* là sự kiện không thể xảy ra với mọi phép thử. Sự kiện này được kí hiệu là $\emptyset$, trùng với tập rỗng.

2.2. Quan hệ giữa các sự kiện

Các sự kiện có thể được liên hệ với nhau qua các phép tính trên tập hợp sau

- **Quan hệ kéo theo**
  Sự kiện A kéo theo sự kiện B, kí hiệu $A \subset B$, xảy ra khi và chỉ khi nếu A xảy ra thì B xảy ra.
- **Quan hệ tương đương**
  Sự kiện A tương đương với sự kiện B khi và chỉ khi $A \subset B$ và $B \subset A$. Kí hiệu $A = B$.
- **Tổng của sự kiện (hoặc hợp của sự kiện)**
  Tổng của hai sự kiện A và B là sự kiện được kí hiệu là $A \cup B$ (hoặc $A + B$). Sự kiện $A \cup B$ xảy ra khi và chỉ khi ít nhất A hoặc B xảy ra
  Tổng của một dãy các sự kiện $A_1, A_2, \ldots, A_n$ là sự kiện $\bigcup_{i=1}^{n} A_i$, xảy ra khi và chỉ khi ít nhất một trong các sự kiện A_i diễn ra.
- **Tích của sự kiện (hoặc giao của sự kiện)**
  Tổng của hai sự kiện A và B là sự kiện được kí hiệu là $A \cap B$ (hoặc AB). Sự kiện $A \cap B$ xảy ra khi và chỉ khi cả A hoặc B xảy ra
  Tổng của một dãy các sự kiện $A_1, A_2, \ldots, A_n$ là sự kiện $\bigcap_{i=1}^{n} A_i$, xảy ra khi và chỉ khi ít nhất tất cả sự kiện A_i diễn ra.
- **Sự kiện xung khắc**
  Hai sự kiện A và B được gọi là xung khắc nếu như sự kiện AB là sự kiện không thể, đồng nghĩa với $A \cap B = \emptyset$.
- **Hệ đầy đủ các sự kiện**
  Dãy các sự kiện $A_1, A_2, \ldots, A_n$ được gọi là một hệ đầy đủ các sự kiện nếu như thỏa mãn các điều sau
  - Các sự kiện xung khắc từng đôi một, hay $A_i \cap A_j = \emptyset$, với mọi i khác j, với $i, j = 1, 2, \ldots, n$.
  - Tổng là sự kiện chắc chắn, nghĩa là $\bigcup_{i=1}^{n} A_i = \Omega$.
- **Sự kiện đối lập**
  Sự kiện đối của A là $\bar{A}$, nghĩa là A xảy ra khi và chỉ khi $\bar{A}$ không diễn ra, và ngược lại. Như vậy, $A \cap \bar{A} = \emptyset$ đồng thời $A \cup \bar{A} = \Omega$.
- **Tính độc lập của sự kiện (independence)**
  Ta nói các sự kiện *A, B* là *độc lập* với nhau nếu như việc xảy ra hay không xảy ra sự kiện này không ảnh hưởng tới việc xảy ra hay không xảy ra tới sự kiện kia.

Tổng quát, dãy các sự kiện A_1, A_2, ..., A_n là độc lập nếu việc xảy ra hay không xảy ra sự kiện này không ảnh hưởng tới việc xảy ra hay không xảy ra của từng sự kiện khác.

2.3. Định nghĩa xác suất

Xác suất của một sự kiện là con số đặc trưng biểu diễn khả năng xuất hiện của sự kiện đó trong quá trình thực hiện phép thử. Dựa vào bản chất của phép thử, ta có thể suy luận về khả năng xuất hiện của sự kiện. Với cách tiếp cận này, ta có *định nghĩa xác suất theo phương pháp cổ điển*. Khi thực hiện lặp lại nhiều lần độc lập một phép thử, ta có thể tính tần suất xuất hiện (số lần xuất hiện) của một sự kiện nào đó. Tần suất thể hiện khả năng xuất hiện của biến cố, với cách tiếp cận này, ta có *định nghĩa xác suất theo phương pháp thống kê*.

2.3.1. Định nghĩa cổ điển về xác suất. Giả sử phép thử C thỏa mãn hai điều kiện sau:

- *Không gian mẫu có một số hữu hạn phần tử.*
- *Các kết quả xảy ra đồng khả năng.*

Khi đó ta định nghĩa xác suất xảy ra sự kiện A là:

$$P(A) = \frac{\text{số kết quả thuận lợi với } A}{\text{số tất cả các kết quả}}.$$

Dưới ngôn ngữ tập hợp, nếu xem A là một tập con của Ω, ta có định nghĩa

$$P(A) = \frac{\text{số phần tử của } A}{\text{số phần tử của } \Omega} = \frac{|A|}{|\Omega|}.$$

Ví dụ 2.2. Xác suất của sự kiện A trong *ví dụ 2.1.* là

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2}{6} = \frac{1}{3}.$$

Điều này mang ý nghĩa xác suất là tỷ lệ để tung được một số bé hơn hoặc bằng 2 là một phần ba.

Ví dụ 2.3. Một công ty cần tuyển 2 nhân viên. Có 6 người nộp đơn trong đó có 4 nữ và 2 nam. Giả sử xác suất mỗi người được trúng tuyển là như nhau. Tính xác suất sự kiện

a) Hai người trúng tuyển là nam.
b) Hai người trúng tuyển là nữ.
c) Có ít nhất một người nam trúng tuyển.

Việc nhìn nhận một bài toán xác suất dưới góc nhìn của một bài toán tập hợp thường sẽ giúp ích rất nhiều trong việc tính toán.

2.3.2. Định nghĩa thống kê về xác suất. Định nghĩa xác suất theo cổ điển tuy trực quan và dễ hiểu. Tuy nhiên khi số các kết quả vô hạn hoặc không đồng khả năng thì cách tính xác suất cổ điển không áp dụng được.

Giả sử phép thử C có thể được thực hiện độc lập lặp lại nhiều lần trong những điều kiện giống hệt nhau. Nếu trong n lần thực hiện phép thử C, sự kiện A xuất hiện $k_n(A)$ lần thì tỉ số

$$f_n(A) = \frac{k_n(A)}{n}$$

được gọi là tần suất xuất hiện của biến cố A trong n phép thử. Người ta chứng minh được (**định lý luật số lớn**) khi n tăng lên vô hạn thì $f_n(A)$ tiến đến một giới hạn xác định. Ta định nghĩa giới hạn này là xác suất của sự kiện A

$$P(A) = \lim_{n\to\infty} f_n(A).$$

Trong thực tế $P(A)$ được tính bằng cách lấy xấp xỉ fn(A) với n đủ lớn. Có thể kiểm chứng điều này bằng việc tính xác suất mặt sấp khi tung đồng xu. Ta biết xác suất này là 1/2 với một đồng xu không bị hư hỏng. Hãy thử tung đồng xu 20 lần và ghi nhận số mặt sấp. Lấy số lượng mắt sấp chia cho 20 ta sẽ được một số rất gần với 1/2.

2.3.3. Định nghĩa xác suất theo hình học. Giả sử không gian mẫu Ω có thể biểu diễn dưới dạng một miền nào đó có diện tích (thể tích, độ dài) hữu hạn và sự kiện A tương ứng với một miền con của Ω thì ta xác suất của sự kiện A được định nghĩa:

$$P(A) = \frac{\text{diện tích (độ dài, thể tích) của } A}{\text{diện tích (độ dài, thể tích) của } \Omega}.$$

Điều này thường được dùng khi ta biểu diễn tập hợp qua **biểu đồ Venn** hoặc dùng trong trường hợp xác suất liên tục.

2.3.4. Các tính chất và định lý xác suất.

- **Tính chất**
  Với mọi sự kiện A, $0 \leq P(A) \leq 1$. Xác suất của *sự kiện không thể* bằng 0, xác suất của *sự kiện chắc chắn* bằng 1, hay $P(\emptyset) = 0$ và $P(\Omega) = 1$.
- **Quy tắc cộng xác suất**
  Với hai sự kiện A và B ta có
  $$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$
- **Xác suất của sự kiện đối**
  Với một sự kiện A, xác suất của sự kiện đối $\bar{A}$ được tính bằng
  $$P(\bar{A}) = 1 - P(A).$$
- **Độc lập xác suất**
  Hai sự kiện A và B độc lập với nhau khi và chỉ khi
  $$P(A \cap B) = P(A)P(B).$$

2.3.5. Xác suất có điều kiện (conditional probability).

- **Định nghĩa**
  Xác suất của sự kiện B được tính trong điều kiện biết rằng đã xảy ra sự kiện A được gọi là xác suất của B với điều kiện A. Kí hiệu là $P(B|A)$. Ví dụ như xác suất trời mưa buổi tối nếu đã biết mưa buổi sáng.
- **Tính chất**
  Với hai sự kiện A và B ta có
  $$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$
  Để ý rằng, nếu A và B độc lập, ta có
  $$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B).$$
  Điều này hợp lý bởi xác suất của B là không phụ thuộc vào việc A có xảy ra hay không.

- **Quy tắc nhân xác suất**
  Công thức trên thường được viết dưới dạng tương đương

$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B).$$

- **Công thức xác suất đầy đủ**
  Với A_1, A_2, ..., A_n là một hệ đầy đủ và một sự kiện B của cùng một phép thử, ta có

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$

- **Công thức Bayes (Bayes' Theorem)**
  Cùng với điều kiện như trên, với mọi sự kiện A_k của hệ và giả sử $P(B) > 0$, ta có

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum\limits_{i=1}^{n} P(A_i)P(B|A_i)}.$$

  Trong thực tế các xác suất $P(A_1)$, $P(A_2)$, ..., $P(A_n)$ đã biết và được gọi là *xác suất tiền nghiệm*. Sau khi quan sát biến cố B xảy ra, các xác suất của A_k được tính dựa trên thông tin này (xác suất có điều kiện $P(A_k|B)$) được gọi là *xác suất hậu nghiệm*. Vì vậy công thức Bayes còn được gọi là *công thức tính xác suất hậu nghiệm*.

3. Biến ngẫu nhiên (Random variable) và các đại lượng liên quan

3.1. Biến ngẫu nhiên

3.1.1. Định nghĩa. *Biến ngẫu nhiên* X là đại lượng nhận các giá trị nào đó phụ thuộc vào các yếu tố ngẫu nhiên.

Ví dụ 3.1. Các đại lượng sau là biến ngẫu nhiên

- Số nốt xuất hiện khi gieo một con xúc sắc
- Tuổi thọ của một thiết bị đang hoạt động
- Số cuộc gọi đến một tổng đài
- Sai số khi đo lường một đại lượng vật lý

3.1.2. Phân loại. Biến ngẫu nhiên được phân ra thành hai loại

- *Biến ngẫu nhiên rời rạc*, nếu chỉ nhận một số hữu hạn hoặc vô hạn đếm được các giá trị. Nghĩa là các giá trị có thể liệt kê thành một dãy x_1, x_2, ...
- *Biến ngẫu nhiên liên tục*, nếu tìm được hai số thực $a < b$ sao cho X nhận được mọi giá trị nằm giữa a và b.

Để phù hợp cho chương trình Toán mô hình, chúng ta sẽ chỉ tạm quan tâm tới biến ngẫu nhiên rời rạc.

3.1.3. Phân phối xác suất của biến ngẫu nhiên rời rạc. Kí hiệu $P(X = x)$ biểu diễn xác suất mà biến ngẫu nhiên X nhận giá trị x trong không gian mẫu Ω. Ta có khái niệm của *bảng phân phối xác suất*. Bảng phân phối xác suất dùng để biểu diễn biến ngẫu nhiên rời rạc, gồm hai dòng, dòng thứ nhất ghi các giá trị của biến ngẫu nhiên, và dòng thứ hai ghi xác suất tương ứng.

X	x_1	x_2	$\ldots$	x_n	$\ldots$
P	p_1	p_2	$\ldots$	p_n	$\ldots$

Bảng 1: Bảng phân phối xác suất

Trong đó, $p_i = P(X = x_i)$, n có thể là vô hạn những phải bảo đảm điều kiện $\sum P(X = x_i) = \sum p_i = 1$.

3.2. *Các đại lượng liên quan*

3.2.1. Giá trị kì vọng. *Kì vọng* hoặc *giá trị trung bình* (average, mean value, expected value) của biến ngẫu nhiên rời rạc X kí hiệu là $E(X)$ và được tính theo công thức

$$E(X) = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i.$$

với $p_i = P(X = x_i)$ tương ứng.

Nếu biến ngẫu nhiên X nhận vô hạn đếm được các giá trị thì $E(X)$ xác định nếu tổng trên trở thành một chuỗi hội tụ tuyệt đối. Ý nghĩa của giá trị kì vọng là trung bình theo xác suất của tất cả các giá trị có thể có của biến ngẫu nhiên.

3.2.2. Phương sai. *Phương sai* (variance) hay độ lệch bình phương trung bình của biến ngẫu nhiên X là đại lượng đo lường sự phân tán bình phương trung bình của X xung quanh giá trị trung bình $E(X)$. Phương sai của X được kí hiệu là $D(x)$ hay var$X$ và được xác định bởi công thức

$$D(X) = E[(X - E(X))^2] = \sum_{i=1}^{n} (x_i - E(X))^2 \cdot p_i.$$

4. Ứng dụng

4.1. *Giá trị kì vọng của trò chơi*

Xét 2 bài toán sau:

i) Nhà cái tổ chức trò chơi quay số trúng thưởng từ 1 tới 100 với thể lệ như sau :
   - Mỗi người chơi phải đóng \$2
   - Nếu người chơi quay trúng số là bội của 10 thì được thưởng \$10
   - Nếu người chơi quay trúng số là bội của 20 thì được thưởng thêm \$10

   Hỏi bên nào sẽ thật sự là người thắng cuộc trong trò chơi này, bên nhà cái hay bên người chơi?

ii) Hai người I và II chơi trò chơi "chẵn và lẻ." Người thứ I là Lẻ và người thứ II là Chẵn. Mỗi lần chơi, hai người sẽ cùng lúc gọi ra số 1 hoặc 2, tổng của hai số thuộc loại nào (chẵn hoặc lẻ) thì người ứng với loại đó là người thắng. Người thua ở mỗi lần chơi sẽ phải trả người thắng số tiền ứng với tổng của hai số hai người gọi trước đó. Hỏi người nào có chiến thuật để "luôn thắng" trong trò chơi này nếu chơi trong một thời gian dài ?

Hai bài toán trên là bài toán điển hình cho ứng dụng của xác suất trong trò chơi cũng như lập mô hình toán học. Bài thứ hai thì phức tạp hơn vì ở đây có hai người chơi cùng lúc nên mỗi người cùng có thể đặt ra "chiến lược tối ưu" (optimal strategy) để thắng, trong khi ở bài thứ nhất thì người chơi chỉ "cầu may." Để thấy tầm quan trọng của xác suất ta sẽ lần lượt giải từng bài.

Lời giải bài i). Xác suất để người chơi thắng \$10 là $\frac{5}{100}$ (quay trúng các số 10,30,50,70,90).
Xác suất để người chơi thắng \$20 là $\frac{5}{100}$ (quay trúng các số 20,40,60,80,100).
Xác suất để người chơi không thắng gì là $\frac{90}{100}$ (các trường hợp còn lại).
Vậy trung bình người chơi sẽ thắng :

$$E(X) = 10 \cdot \frac{5}{100} + 20 \cdot \frac{5}{100} + 0 \cdot \frac{90}{100} = \frac{3}{2}.$$

Tuy nhiên do người chơi vẫn phải đóng \$2 để chơi nên số tiền trung bình người chơi có thể đạt được chỉ là

$$\frac{3}{2} - 2 = -\frac{1}{2}.$$

Như vậy nếu chơi lâu dài thì bên nhà cái luôn được lợi hơn (hiển nhiên): trung bình mỗi người chơi thì họ thu lợi được một nửa dollar. ☐

Lời giải bài ii). Mỗi lượt chơi chỉ có hoặc người I hoặc người II sẽ thắng, nên ta mặc định "chiến lược" ở đây là cách người I và người II sẽ gọi số theo tần suất như thế nào để đảm bảo mình sẽ "có thể" luôn thắng (vì chơi về lâu về dài).

Gọi p là tần suất người I gọi số 1 và q là tần suất người II gọi số 1. Giá trị kì vọng số tiền người I sẽ nhận được là :

$$\begin{aligned} E_1(X) &= 3[p(1-q) + q(1-p)] - 2pq - 4(1-p)(1-q) \\ &= 7p + 7q - 12pq - 4. \end{aligned}$$

Ngược lại số tiền người II sẽ thắng (hoặc thua) bằng số tiền người I mất đi (hoặc thắng) nên ta có giá trị kì vọng của người II:

$$E_2(X) = 12pq + 4 - 7p - 7q.$$

Chúng ta cần phải tìm liệu xem có tồn tại $p \in [0,1]$ sao cho với mọi $q \in [0,1]$ thì $E_1(X)$ luôn không âm, hoặc ngược lại tồn tại $q \in [0,1]$ sao cho với mọi $p \in [0,1]$ thì $E_2(X)$ luôn không âm. Ta xét hai trường hợp

- *Trường hợp 1:* $E_1(X)$ không âm.
  Ta có

  $$E_1(X) = 7p + 7q - 12pq - 4 = q(7 - 12p) + 7p - 4 \geq 0, \text{ với mọi } q \in [0,1].$$

  Vì vế trái là một đại lượng tuyến tính theo q, ta chỉ cần xét giá trị của q ở biên, tức $q = 0$ và $q = 1$. Bằng việc thế trực tiếp, $q = 0$ sẽ cho ta $p \geq 4/7$ và $q = 1$ sẽ cho ta $p \leq 3/5$.
  Như vậy với p thỏa mãn $4/7 \leq p \leq 3/5$ thì trung bình người I luôn thắng hơn so với người II.
- *Trường hợp 2:* $E_2(X)$ không âm.
  Bằng việc làm tương tự như trên, ta sẽ thấy không tồn tại q để thỏa mãn điều kiện này.

Tất nhiên với nếu số lượt chơi là ít thì không hẳn nhiên lúc nào người I cũng thắng, nhưng nếu chơi đủ lâu thì với chiến lược trên, người I sẽ luôn trên cơ so vối người II (chiến lược ra số 1 với tỉ lệ nằm giữa 4/7 và 3/5). ☐

4.2. Chuỗi Markov (Markov chain)

Chuỗi Markov là một chuỗi các biến ngẫu nhiên $X_1, X_2, X_3, \ldots,$ có *thuộc tính Markov*, tức là việc xác định (dự đoán) phân bố xác suất có điều kiện của X_{n+1} khi cho biết các trạng thái quá khứ là một hàm chỉ phụ thuộc X_n.

$$P(X_{n+1} = x \mid X_n, \ldots, X_2, X_1) = P(X_{n+1} = x \mid X_n).$$

Tập tất cả các giá trị có thể có của các biến ngẫu nhiên X_i được gọi là *không gian trạng thái* S, x là một trạng thái nào đó trong không gian trạng thái S.

4.2.1. Chuỗi markov rời rạc thuần nhất. Là chuỗi Markov thỏa mãn hàm xác suất $P(X_{n+1} = x \mid X_n)$ là như nhau với mọi n. Khi đó X_n có tập các kết quả là như nhau. Ta thường sử dụng **Ma trận xác suất chuyển** để tính xác suất $P(X_n = x)$ tại thời điểm n như sau.

Đặt $p_{ij} = P(X_{n+1} = i \mid X_n = j)$ không phụ thuộc vào n. p_{ij} là xác suất để từ trạng thái i sau một bước sẽ chuyển qua trạng thái j. Đặt ${p_{ij}}^{(k)} = P(X_{n+k} = i \mid X_n = j)$.

Ta có ma trận vuông $P = (p_{ij})$ được gọi là *ma trận xác suất chuyển sau một bước*. $P^{(k)} = (p_{ij}^{(k)})$ được gọi là *ma trận xác suất chuyển sau k bước*.

Ta có một hệ quả quan trọng gọi là **Phương trình Chapman-Kolmogorov**. Với định nghĩa P và $P^{(n)}$ như trên, với mọi $n \geq 0$, ta có

$$P^{(n+1)} = P^{(n)} \cdot P.$$

Đặt $p_i^{(n)} = P(X_n = i)$, $n = 0, 1, 2, \ldots$ Ma trận hàng $Q^{(n)} = (p_i^{(n)})$ được gọi là *phân bố của hệ tại thời điểm n*. Ta chứng minh được

- $Q^{(n+1)} = Q^{(n)} \cdot P$
- $Q^{(n+m)} = Q^{(n)} \cdot P^{(m)}$
- $Q^{(n+1)} = Q \cdot P^{(n)}$

Từ Chapman-Kolmogorov, ta suy ra được ba phương trình trên là như nhau.

Như vậy ta đã hệ thống hóa được chuỗi Markov rời rạc thuần nhất qua các bộ $(X(n), P, Q)$ với $X(n)$ là chuỗi các biến ngẫu nhiên rời rạc $X_1, X_2, \ldots, X_n$.

4.2.2. Ví dụ. Cho ba thành phố *A, B, C* có số dân hiện nay lần lượt là 2.5, 4, 5.5 triệu người. Thống kê cho thấy rằng mỗi năm ở thành phố A trung bình có 10% số dân hiện tại sẽ qua B và 15% số dân hiện tại sẽ qua C để kiếm việc làm, riêng thành phố B trung bình có 25% số dân sẽ qua A và 15% sẽ qua C, thành phố C có trung bình 30% số dân hiện tại sẽ qua A nhưng sẽ không ai qua B. Hỏi

a) Trong vòng 5 năm tới thì tổng số dân ở hai thành phố A và B là bao nhiêu?
b) Lập mô hình tính số dân của thành phố *A, B, C* trong n năm tính từ năm nhất.
c) Lập lại mô hình ở *b*) nếu như mỗi thành phố có tỉ lệ sinh thêm con mỗi năm là 10% dân số, biết rằng % số dân chuyển đi là không tính những trẻ em được sinh thêm.

5. Tham khảo

[1] *Lý thuyết Xác suất và Thống kê Toán*, Tô Anh Dũng, NXB ĐHQG TP.HCM.

[2] Sách hướng dẫn học tập Lý thuyết Xác suất và Thống kê Toán – Ts. Lê Bá Long http://www.e-ptit.edu.vn/hoctap/hoclieu/XSTK.pdf

[3] *Expected value and the Game of Craps*, Blake Thornton.

[4] http://www.dehn.wustl.edu/~blake/courses/WU-Ed6021-2012-Summer/handouts/Expected_Value.pdf

[5] http://www.mathguide.com/activities2/ExpectedValueGames.pdf

[6] https://www.math.ucla.edu/~tom/Game_Theory/mat.pdf

[7] https://en.wikipedia.org/wiki/Markov_chain