TỐNG ĐÌNH QUỲ

GIÁO TRÌNH

XÁC SUẤT THỐNG KÊ

(Tái bản lần thứ năm)

NHÀ XUẤT BẢN BÁCH KHOA – HÀ NỘI

LỜI NÓI ĐẦU

Lý thuyết xác suất và thống kê toán học là một ngành khoa học đang giữ vị trí quan trọng trong các lĩnh vực ứng dụng rộng rãi và phong phú của đời sống con người. Cùng với sự phát triển mạnh mẽ của khoa học và công nghệ, nhu cầu hiểu biết và sử dụng các công cụ ngẫu nhiên trong phân tích và xử lý thông tin ngày càng trở nên đặc biệt cần thiết. Các kiến thức và phương pháp của xác suất và thống kê đã hỗ trợ hữu hiệu các nhà nghiên cứu trong nhiều lĩnh vực khoa học khác nhau như vật lý, hóa học, sinh y học, nông học, kinh tế học, xã hội học, ngôn ngữ học...

Trong một chục năm gần đây, giáo trình xác suất thông kê đã trở thành cơ sở của nhiều ngành học trong các trường đại học và cao đẳng, từ đó xuất hiện nhu cầu học tập và nghiên cứu ứng dụng rất lớn, nhất là đối với sinh viên các ngành khoa học không chuyên về toán. Để thoả mãn yêu cầu đó, giáo trình này cố gắng đáp ứng đòi hỏi của đông đảo sinh viên nhằm hiểu biết sâu sắc hơn các khái niệm và phương pháp tính xác suất và thông kê để học tập đạt hiệu quả cao hơn cũng như ứng dụng môn học vào ngành học và môn học khác.

Giáo trình xác suất thống kê được viết cho thời gian giảng dạy là 60 tiết học. Do đối tượng sinh viên rất đa dạng với trình độ toán cơ bản khác nhau, chúng tôi đã cố gắng tìm những cách tiếp cận đơn giản và hợp lý, và như vậy đã buộc phải bớt đi phần nào sự chặt chẽ hình thức (vốn rất đặc trưng cho toán học) để giúp bạn đọc tiếp cận dễ dàng hơn bản chất xác suất của các vấn đề đặt ra và tăng cường kỹ năng phân tích, xử lý các tình huống, từ đó dần dần hình thành một hệ thống khái niệm khá đầy đủ để đi sâu giải quyết các bài toán ngày càng phức tạp hơn.

Giáo trình được chia thành 6 chương gồm 3 chương dành cho phần xác suất và 3 chương cho phần phân tích thống kê. Những khái niệm và công thức cơ bản được trình bày tương đối đơn giản, dễ hiểu và được minh hoạ bằng nhiều thí dụ áp dụng. Các chứng minh khó được lượt bớt có chọn lọc để giáo trình không quá cồng kềnh, mặc dù vậy các công thức và vấn đề liên quan đều được nhắc đến đầy đủ để tiện không chỉ cho học tập sâu hơn, mà còn có ích cho những bạn đọc muốn tra cứu, tìm tòi phục vụ cho ứng dụng và tính toán thống kê. Cuối mỗi chương có một loạt bài tập dành để bạn đọc tự giải nhằm hiểu biết sâu sắc hơn lý thuyết và rèn luyện kỹ năng thực hành.

Hy vọng rằng giáo trình có ích cho bạn đọc xa gần, các sinh viên, cán bộ giảng dạy ở các trường đại học và cao đẳng, các cán bộ khoa học và kinh tế muốn tự học và tự nghiên cứu xác suất thống kê – môn học thường được coi là khó tiếp thu. Tác giả cũng cám ơn mọi ý kiến góp ý để quyển sách sẽ ngày càng được hoàn thiện hơn để góp phần nâng cao chất lượng dạy và học môn học này.

Trong lần tái bản này tại Nhà xuất bản Bách Khoa – Hà Nội, một số lỗi chế bản đã được sửa chữa. Tác giả một lần nữa tỏ lời cảm ơn đến những ý kiến góp ý của đông đảo bạn đọc để cải tiến giáo trình trong lần tái bản tiếp theo.

TÁC GIẢ

Chương I

SỰ KIỆN NGẪU NHIÊN VÀ PHÉP TÍNH XÁC SUẤT

§1. KHÁI NIỆM MỞ ĐẦU

1.1. Sự kiện ngẫu nhiên

Khái niệm thường gặp trong lý thuyết xác suất là *sự kiện* (mà không thể định nghĩa chặt chẽ). Sự kiện được hiểu như là một sự việc, một hiện tượng nào đó của cuộc sống tự nhiên và xã hội.

Khi thực hiện một tập hợp điều kiện xác định, nói tắt là bộ điều kiện, gọi là một *phép thử*, có thể có nhiều kết cục khác nhau.

Thí dụ 1.1. Gieo một con xúc sắc đồng chất trên một mặt phẳng (phép thử). Phép thử này có 6 kết cục là: xuất hiện mặt 1, mặt 2,..., mặt 6 chấm. Mỗi kết cục này cùng với các kết quả phức tạp hơn như: xuất hiện mặt có số chấm chẵn, mặt có số chấm bội 3, đều có thể coi là các sự kiện.

Như vậy kết cục của một phép thử là một trường hợp riêng của sự kiện. Để cho tiện lợi sau này, ta ký hiệu sự kiện bằng các chữ cái in hoa A, B, C, ... Sự kiện được gọi là *tất yếu*, nếu nó chắc chắn xảy ra, và được gọi là *bất khả*, nếu nó không thể xảy ra khi thực hiện phép thử. Còn nếu sự kiện có thể xảy ra hoặc không sẽ được gọi là *sự kiện ngẫu nhiên*. Từ đó, theo một nghĩa nào đó, có thể coi các sự kiện tất yếu, ký hiệu là U, và bất khả, ký hiệu là V, như các trường hợp riêng của sự kiện ngẫu nhiên. Thí dụ, dưới những điều kiện xác định, nước đóng băng ở 0^0C là sự kiện tất yếu; khi gieo một con xúc xắc, việc xuất hiện mặt bảy chấm là sự kiện bất khả...

Để mô tả một phép thử người ta xác định tập hợp các kết cục có thể có. Tập hợp tất cả các kết cục của một phép thử (được gọi là các ***sự kiện sơ cấp***, ký hiệu là ω_i) tạo thành không gian các sự kiện sơ cấp, ký hiệu là $\Omega = \{\omega_i, i \in I\}$, I là tập chỉ số, có thể vô hạn (đếm được hoặc không đếm được). Dễ thấy trong thí dụ 1.1, nếu ký hiệu A_i – sự kiện xuất hiện mặt i chấm ($i = \overline{1,6}$) thì $\Omega = \{A_1, A_2, A_3, A_4, A_5, A_6\} = \{A_i, i = \overline{1,6}\}$.

Trong nhiều hiện tượng hàng loạt khi thực hiện nhiều lần cùng một phép thử, ta thấy tần suất xuất hiện một sự kiện A nào đó chênh lệch không nhiều so với một số đặc trưng cho khả năng xuất hiện A. Số đó được gọi là *xác suất* xuất hiện A và được ký hiệu là $P(A)$. Như vậy nếu viết $P(A) = p$ có nghĩa là xác suất xảy ra sự kiện A là bằng p.

Một câu hỏi tự nhiên là. Do đâu có sự kiện ngẫu nhiên? Và chúng ta có thể nhận biết được chúng không? Thực ra mỗi sự kiện đều xảy ra theo quy luật nào đó; song do điều kiện thiếu tri thức, thông tin và phương tiện cần thiết (cả về kinh phí, thiết bị lẫn thời gian) nên ta không có khả năng nhận thức đầy đủ về sự kiện đó. Vấn đề càng trở nên khó khăn hơn khi chỉ cần có một sự thay đổi bất ngờ dù rất nhỏ của bộ điều kiện đã làm thay đổi kết cục của phép thử. Cho nên bài toán xác định bản chất xác suất của một sự kiện bất kỳ trong một phép thử tùy ý là không thể giải được.

1.2. Phép toán và quan hệ của các sự kiện

Về mặt toán học, việc nghiên cứu quan hệ và phép toán trên tập các sự kiện cho phép ta xác định chúng thực chất hơn.

(i) *Tổng* của A và B, ký hiệu là $A + B$, chỉ sự kiện khi có xuất hiện ít nhất một trong hai sự kiện trên.

(ii) *Tích* của A và B, ký hiệu là AB, chỉ sự kiện khi có xuất hiện đồng thời cả hai sự kiện trên.

(iii) *Đối lập* của A, ký hiệu là $\overline{A}$, chỉ sự kiện không xuất hiện A. Rõ ràng đối lập có tính tương hỗ $\overline{\overline{A}} = A$ và $A + \overline{A} = U$, $A\overline{A} = V$, $\overline{U} = V$.

(iv) *Xung khắc:* hai sự kiện A và B được gọi là xung khắc nếu chúng không thể đồng thời xảy ra, tức là $AB = V$.

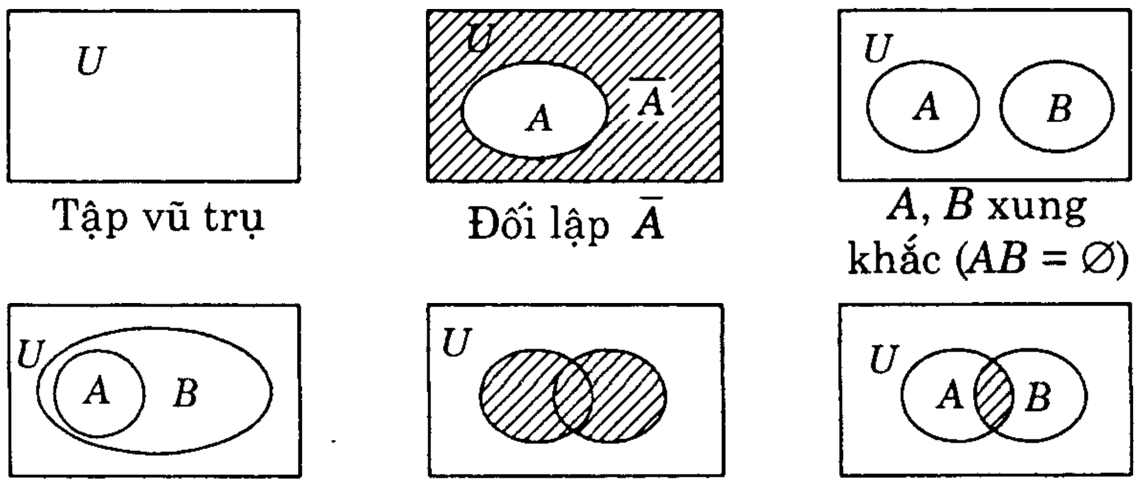
(v) *Kéo theo*, ký hiệu $A \Rightarrow B$, chỉ nếu xuất hiện A thì xuất hiện B.

(vi) *Tương đương*, ký hiệu $A = B$, chỉ việc nếu xuất hiện A thì xuất hiện B và ngược lại.

(vii) *Hiệu* của A và B, ký hiệu $A - B$ (hoặc $A\backslash B$), chỉ sự kiện xuất hiện A nhưng không xuất hiện B, tức là $A - B = A\overline{B}$.

Các khái niệm cho thấy tính đối lập, tổng, tích và hiệu của hai kiện tương ứng với bù, hợp, giao và hiệu của hai tập hợp. Như vậy có thể sử dụng các tính chất của các phép toán trên tập hợp cho các phép toán trên sự kiện, chẳng hạn dùng sơ đồ Ven trong thí dụ sau đây.

Thí dụ 1.2. Ký hiệu U là tập vũ trụ, V là tập $\varnothing$ (rỗng). Khi đó A và B sẽ là các tập con của U và các phép toán trên A và B có thể minh họa bằng sơ đồ Ven (xem hình 1.1).



Tập vũ trụ | Đối lập $\overline{A}$ | A, B xung khắc ($AB = \varnothing$)

Kéo theo $A \Rightarrow B$ | Tổng $A + B$ | Tích AB

Hình 1.1

Từ đó, dễ dàng chỉ ra các công thức sau:

$A + B = B + A$, $AB = BA$ (giao hoán);

$A + (B + C) = (A + B) + C$, $A(BC) = (AB)C$ (kết hợp);

$A(B + C) = AB + AC$ (phân phối);

$A + U = U, A + V = A, A + A = A$;

$AU = A, AV = V, AA = A$.

Thí dụ 1.3. Chọn từ một lô hàng ra 5 sản phẩm và ta quan tâm đến số phế phẩm trong 5 sản phẩm đó (phép thử).

a) Xác định các sự kiện sơ cấp.

b) Biểu diễn các sự kiện sau theo các sự kiện sơ cấp: có nhiều nhất 1 phế phẩm; có không quá 4 phế phẩm, có ít nhất 1 phế phẩm.

Giải. a) Ký hiệu A_i – trong 5 sản phẩm có i phế phẩm. Rõ ràng $i = \overline{0,5}$ và $\Omega = \{A_0, A_1, A_2, A_3, A_4, A_5\}$.

b) Gọi A, B và C là các sự kiện tương ứng. Dễ dàng biểu diễn $A = A_0 + A_1$, $B = A_0 + A_1 + A_2 + A_3 + A_4 = \overline{A_5}$, $C = A_1 + A_2 + A_3 + A_4 + A_5 = \overline{A_0}$.

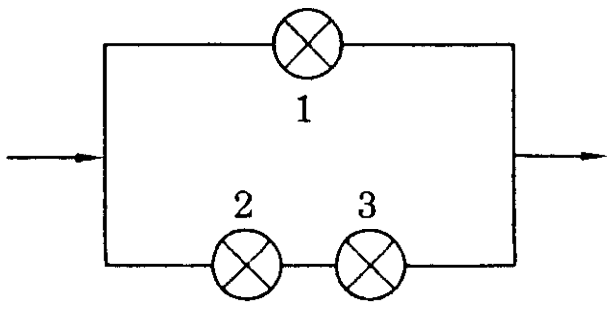
Thí dụ 1.4. Cho sơ đồ mạng điện trên hình 1.2 gồm 3 bóng đèn. Việc mạng mất điện (sự kiện A) chỉ có thể xảy ra do cháy các bóng đèn (ký hiệu là A_1, A_2, A_3). Hãy biểu diễn A theo các A_i, $i = 1, 2, 3$).

Giải. A xuất hiện khi xảy ra một trong 3 trường hợp:

(i) cả ba bóng cháy,

(ii) cháy hai bóng 1 và 2,

(iii) cháy hai bóng 1 và 3.



Hình 1.2

Từ đó ta có $A = A_1A_2A_3 + A_1A_2\overline{A_3} + A_1\overline{A_2}A_3$.

Có thể dùng tính chất của mạng song song và nối tiếp để có một biểu diễn khác gọn hơn:

$$A = A_1(A_2 + A_3).$$

Trong nhiều bài tập, việc xác định số lượng các sự kiện sơ cấp đưa đến sử dụng các kết quả của lý thuyết tổ hợp.

1.3. Giải tích kết hợp

Việc đếm số các kết cục của một phép thử dựa vào mô hình: chọn hú họa ra k phần tử từ n phần tử cho trước. Nếu phân biệt thứ tự các phần tử chọn ra, ta có khái niệm chỉnh hợp; nếu thứ tự không phân biệt, ta có tổ hợp.

(i) *Chỉnh hợp:* chỉnh hợp chập k từ n là một nhóm có thứ tự gồm k phần tử lấy từ n đã cho. Đó chính là một nhóm gồm k phần tử khác nhau được xếp theo thứ tự nhất định. Số các chỉnh hợp như vậy, ký hiệu là ($k \le n$).

$$A_n^k = n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}. \qquad (1.1)$$

(ii) *Chỉnh hợp lặp:* chỉnh hợp lặp chập k từ n là một nhóm có thứ tự gồm k phần tử có thể giống nhau lấy từ n đã cho. Đó chính là một nhóm gồm k phần tử có thể lặp lại và được xếp theo thứ tự nhất định. Số các chỉnh hợp lặp như vậy, ký hiệu là

$$\widetilde{A}_n^k = n^k. \qquad (1.2)$$

(iii) *Hoán vị:* hoán vị của n là một nhóm gồm n phần tử được sắp xếp theo một thứ tự nào đó. Rõ ràng số các hoán vị như vậy, ký hiệu là P_n, chính là số các chỉnh hợp A_n^n và

$$P_n = n! \qquad (1.3)$$

(iv) *Tổ hợp:* tổ hợp chập k từ n là một nhóm (không phân biệt thứ tự) gồm k phần tử khác nhau lấy từ n đã cho. Số các tổ hợp như vậy, ký hiệu là ($k \le n$)

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}. \quad (1.4)$$

Thí dụ 1.5. Cho một tập hợp gồm 3 phần tử $\{a, b, c\}$. Có thể tạo ra bao nhiêu nhóm gồm 2 phần tử chọn từ tập trên?

Giải:

(i) Nếu ta để ý đến thứ tự các phần tử và mỗi phần tử chỉ được chọn một lần, số nhóm thu được sẽ là $A_3^2 = 3.2 = 6$; đó là $\{a, b\}$; $\{b, a\}$; $\{a, c\}$; $\{c, a\}$; $\{b, c\}$, $\{c, b\}$.

(ii) Nếu vẫn để ý đến thứ tự, nhưng mỗi phần tử được chọn nhiều lần, số nhóm thu được trở thành $\widetilde{A}_3^2 = 3^2 = 9$; đó là:

$\{a, b\}$; $\{b, a\}$; $\{a, c\}$; $\{c, a\}$; $\{b, c\}$, $\{c, b\}$; $\{a, a\}$; $\{b, b\}$; $\{c, c\}$.

(iii) Nếu không để ý đến thứ tự các phần tử và chúng chỉ được chọn một lần, số nhóm thu được trở thành $C_3^2 = 3$; đó là

$\{a, b\}$; $\{a, c\}$; $\{b, c\}$.

Thí dụ 1.6. Một lớp phải học 6 môn trong học kỳ, mỗi ngày học 3 môn. Hỏi có bao nhiêu cách xếp thời khóa biểu trong 1 ngày?

Giải. Số cách xếp cần tìm chính là số cách ghép 3 môn từ 6 môn, trong đó các cách ghép sẽ khác nhau nếu có ít nhất một môn khác nhau hoặc thứ tự môn khác nhau. Từ đó theo (1.1) ta có số cách cần tìm là $A_6^3 = 6.5.4 = 120$.

Thí dụ 1.7. Có thể đánh số được bao nhiêu xe nếu chỉ dùng 3 con số từ 1 đến 5?

Giải. Mỗi số thứ tự của một xe dễ thấy là chỉnh hợp lặp chập 3 từ 5. Từ đó theo (1.2) ta có số lượng xe được đánh số sẽ là

$$\widetilde{A}_5^3 = 5^3 = 125.$$

Thí dụ 1.8. Có bao nhiêu cách lập một hội đồng gồm 3 người chọn trong số 8 người?

Giải. Hội đồng là một nhóm 3 người lấy từ 8 người, do đó theo (1.4) sẽ có $C_8^3 = 8!/(3!5!) = 56$ cách lập.

Cuối cùng, để ý là ta đã rất quen thuộc với khái niệm tổ hợp được dùng trong công thức nhị thức Niu-tơn

$$(x+a)^n = C_n^0 x^n + C_n^1 x^{n-1} a + \ldots + C_n^k x^{n-k} a^k + \ldots + C_n^n a^n.$$

Từ đó có thể dễ dàng chứng minh (để ý $C_n^0 = C_n^n = 1$)

$$C_n^k = C_n^{n-k},\; C_n^k = C_{n-1}^{k-1} + C_{n-1}^k.$$

§2. CÁC ĐỊNH NGHĨA CỦA XÁC SUẤT

2.1. Định nghĩa cổ điển

Trong mục này ta làm việc với các phép thử có kết cục *đồng khả năng*. Khái niệm đồng khả năng đóng vai trò chủ đạo và khó có thể định nghĩa một cách hình thức. Xét thí dụ đơn giản sau đây:

Thí dụ 2.1. Trong một hộp có n viên bi giống nhau về kích cỡ và chỉ khác nhau về màu sắc, trong đó có m bi trắng và $n - m$ bi đỏ. Rút hú họa ra một viên bi (phép thử). Do số viên bi là n nên tổng số các kết cục khác nhau sẽ là n, và vì tính giống nhau của chúng nên mỗi viên bi có cùng khả năng được rút. Bây giờ nếu gọi A là sự kiện rút được bi trắng thì trong số n kết cục đồng khả năng có m kết cục thuận lợi cho A. Vì vậy trực giác cho thấy nên chọn tỷ số m/n làm xác suất của việc xuất hiện A.

Định nghĩa. Cho một phép thử với n kết cục đồng khả năng, trong đó có m kết cục thuận lợi cho A, khi đó

$$P(A) = \frac{m}{n} = \frac{\text{số kết cục thuận lợi cho } A}{\text{tổng số kết cục có thể}}. \tag{2.1}$$

Định nghĩa trên được gọi là *định nghĩa cổ điển* của xác suất. Cách tính xác suất theo (2.1) có ưu điểm là tương đối đơn giản và trực quan, tuy nhiên phạm vi áp dụng rất hạn chế chỉ cho các loại phép thử gồm hữu hạn kết cục đồng khả năng. Trong tính toán thường sử dụng các kết quả (1.1) – (1.4).

Thí dụ 2.2. Gieo đồng thời 2 con xúc sắc giống nhau. Tính xác suất để tổng số chấm thu được bằng 6.

Giải. Phép thử có 6.6 = 36 kết cục (sự kiện sơ cấp) khác nhau đồng khả năng. Gọi A là sự kiện "tổng số chấm bằng 6", thì có tất cả 5 kết cục thuận lợi cho A là $\{1,5\}$, $\{2,4\}$, $\{3,3\}$, $\{4,2\}$ và $\{5,1\}$ (số thứ nhất chỉ số chấm của con xúc sắc 1, số thứ 2 – số chấm của con xúc sắc 2). Vậy $P(A) = 5/36$.

Thí dụ 2.3. Trong hộp có 4 viên bi trắng và 6 viên bi đỏ cùng kích cỡ. Rút hú họa ra 2 bi, tính các xác suất để trong đó có:

a) hai viên trắng;

b) ít nhất 1 viên đỏ;

c) viên thứ hai đỏ.

Giải. Ta dùng định nghĩa cổ điển ở trên.

a) Tổng số cách để rút ra 2 bi có quan tâm đến thứ tự là $A_{10}^2 = 10.9 = 90$, trong đó số cách thuận lợi cho A – rút được 2 bi trắng – là $A_4^2 = 4.3 = 12$; vậy xác suất cần tìm $P(A) = 12/90 = 2/15$. Có thể sử dụng khái niệm tổ hợp để tính xác suất: tổng số cách lấy ra 2 bi từ 10 viên bi là C_{10}^2 (không quan tâm đến thứ tự), trong đó để rút ra 2 bi trắng có C_4^2 cách. Từ đó ta có cùng kết quả như trên.

b) Có thể tính trực tiếp xác suất của B – sự kiện rút được ít nhất 1 bi đỏ (tức là hoặc được 1 hoặc cả 2 bi đỏ). Dễ thấy sự kiện đối lập $\bar{B}$ – cả 2 bi đều trắng – đã có xác suất hiện bằng 2/15. Từ đó $P(B) = 1 - P(\bar{B}) = 13/15$ (xem tính chất của xác suất ngay dưới đây).

c) Gọi C là sự kiện viên bi thứ hai màu đỏ. Số cách thuận lợi cho C bao gồm (có quan tâm đến thứ tự): 6.5 = 30 cách đối với trường hợp viên bi đầu màu đỏ và 4.6 = 24 cách đối với trường hợp bi đầu màu trắng. Từ đó $P(C) = (30 + 24)/90 = 3/5$. Có thể lý luận đơn giản hơn như sau: do viên bi đầu không biết màu sắc nên thông tin về tỷ lệ màu không thay đổi với viên bi thứ hai. Vậy sự kiện C sẽ có cùng xác suất với việc rút hú họa ra 1 bi đỏ từ hộp 10 viên ban đầu và xác suất của sự kiện đó rất dễ tính là 6/10 = 3/5.

Dùng công thức (2.1) dễ dàng chứng minh các tính chất sau đây của xác suất (đúng cho cả các trường hợp định nghĩa khác):

(i) $1 \geq P(A) \geq 0$;

(ii) $P(U) = 1$; $P(V) = 0$;

(iii) Nếu A, B xung khắc thì $P(A + B) = P(A) + P(B)$;

(iv) $P(\bar{A}) = 1 - P(A)$;

(v) Nếu $A \Rightarrow B$ thì $P(A) \leq P(B)$.

Để khắc phục hạn chế của (2.1) chỉ áp dụng cho các phép thử có hữu hạn kết cục, người ta đưa ra *định nghĩa hình học* của xác suất. Gải sử tập hợp (vô hạn) các kết cục đồng khả năng của một phép thử có thể biểu thị bởi một miền hình học G (chẳng hạn đoạn thẳng, một miền mặt cong hoặc khối không gian...), còn tập các kết cục thuận lợi cho A bởi một miền con nào đó $S \subseteq G$. Sẽ rất hợp lý nếu ta định nghĩa xác suất bằng tỷ số độ đo của S với G (phụ thuộc vào S và G mà độ đo có thể là độ dài, diện tích hoặc thể tích...). Như vậy ta có $P(A)$ bằng xác suất để điểm gieo rơi vào S, với giả thiết nó có thể rơi đồng khả năng vào các điểm của G và

$$P(A) = \frac{\text{độ đo } S}{\text{độ đo } G}. \qquad (2.2)$$

Khái niệm "rơi đồng khả năng vào G" có nghĩa là điểm gieo có thể rơi vào bất kỳ điểm nào của G và xác suất để nó rơi vào một miền con nào đó của G tỷ lệ với độ đo của miền ấy, mà không phụ thuộc vào vị trí và hình dạng của miền.

Thí dụ 2.4. Đường dây điện thoại ngầm nối một tổng đài với một trạm dài 1km. Tính xác suất để dây đứt tại nơi cách tổng đài không quá 100m.

Giải. Rõ ràng nếu dây điện thoại đồng chất, khả năng nó bị đứt tại một điểm bất kỳ là như nhau, nên tập hợp các kết cục đồng khả năng có thể biểu thị bằng đoạn thẳng nối tổng đài với trạm. Các kết cục thuận lợi cho A – sự kiện chỗ đứt cách tổng đài không quá 100m – được biểu thị bằng đoạn thẳng có độ dài 100m. Từ đó theo (2.2) $P(A) = 100/1000 = 0,1$.

Một số bài toán thực tế khác có thể đưa về mô hình dạng trên. Chú ý rằng theo cách định nghĩa này thì sự kiện có xác suất bằng 0 vẫn có thể xảy ra (chẳng hạn mũi tên bắn trúng một điểm cho trước...). Tính chất này rất đặc trưng cho các biến ngẫu nhiên liên tục sẽ nghiên cứu ở chương II.

2.2. Định nghĩa thống kê

Điều kiện đồng khả năng của các kết cục một phép thử không phải lúc nào cũng được bảo đảm. Có nhiều hiện tượng xảy ra không theo các yêu cầu của định nghĩa cổ điển, chẳng hạn khi tính xác suất một đứa trẻ sắp sinh là con trai, ngày mai trời mưa vào lúc chính ngọ, v.v...

Có một cách khác để xác định xác suất của một sự kiện. Giả sử tiến hành một loạt n_1 phép thử cùng loại, nếu sự kiện A nào đó xuất hiện trong m_1 phép thử thì ta gọi m_1/n_1 là tần suất xuất hiện A trong loạt phép thử đã cho. Tương tự với loại phép thử thứ hai, thứ ba... ta có các tần suất tương ứng m_2/n_2, m_3/n_3,...

Trên cơ sở quan sát lâu dài các thí nghiệm khác nhau người ta nhận thấy tần suất xuất hiện một sự kiện có tính ổn định, thay đổi rất ít trong các loạt phép thử khác nhau và dao động xung quanh một hằng số xác định. Sự khác biệt đó càng ít khi số phép thử tăng nhiều lên. Hơn nữa đối với các phép thử xét ở mục 2.1 hằng số xác định đó trùng với xác suất theo định nghĩa cổ điển. Đặc tính ổn định của tần suất khi số phép thử tăng lên khá lớn cho phép ta định nghĩa xác suất của sự kiện là trị số ổn định đó của tần suất xuất hiện sự kiện. Nhưng do hằng số đó chưa biết, nên người ta lấy ngay tần suất khi số phép thử đủ lớn làm xác suất của sự kiện. Cách hiểu như vậy được gọi là *định nghĩa thống kê* của xác suất.

Như vậy xác suất ở đây là một giá trị gần đúng và nhiều người cho rằng đó không phải là một định nghĩa thật sự. Tuy nhiên, trong nhiều ngành khoa học thực nghiệm xác suất được xác định theo cách này đạt độ chính xác khá lớn và rất phù hợp với thực tế cũng như với tính toán lý thuyết, nhiều khi sai số phạm phải bé hơn nhiều so với sai số đo của thí nghiệm. Vì thế định nghĩa thống kê vẫn được thừa nhận rộng rãi và rất có ý nghĩa. Ta có thể định nghĩa chặt chẽ hơn về mặt toán học như sau: xác suất của sự kiện là giới hạn của tần suất xuất hiện sự kiện đó khi số phép thử tăng vô hạn. Sự hợp lý của định nghĩa được minh chứng không chỉ bằng thực nghiệm mà cả bằng lý thuyết (sau này ta sẽ thấy rõ trong luật số lớn Béc-nu-li).

Có nhiều thí dụ minh họa tính ổn định của tần suất khi số phép thử khá lớn. Ta có thể tham khảo dưới đây các tần suất xuất hiện mặt sấp khi gieo một đồng tiền nhiều lần:

Người thí nghiệm	***Số lần gieo***	***Số lần sấp***	***Tần suất***
Buýt-phông	4040	2048	0,5080
Piếc-xơn	12000	6019	0,5016
Piếc-xơn	24000	12012	0,5005

Một thí dụ khác: có thể cho rằng xác suất phân rã của một nguyên tử Ra^{226} sau 100 năm là 0,04184 (với độ chính xác tới 5 chữ số sau dấu phẩy); ở đây số lượng nguyên tử tham gia thí nghiệm rất lớn (cỡ $10^{23} - 10^{24}$).

Có thể kiểm tra được rằng xác suất định nghĩa theo thống kê thỏa mãn các tính chất trình bày ở mục trước. Chú ý là trong định nghĩa phải có điều kiện các phép thử lặp lại như nhau, điều này trên thực tế không dễ bảo đảm nên tần suất có thể phụ thuộc vào thời gian. Mặc dù vậy phương pháp xác định xác suất theo tần suất có phạm vi ứng dụng rất lớn trong nhiều ngành khoa học và kỹ thuật. Mặt khác, điểm xuất phát để xây dựng lý thuyết xác suất như là một khoa học cũng chính là việc quan sát tính ổn định thống kê của các tần suất của vô vàn các hiện tượng thực tế. Từ đó dễ hiểu vì sao có thể định nghĩa lý thuyết xác suất như là một khoa học nghiên cứu các mô hình toán học của các hiện tượng ngẫu nhiên có tần suất ổn định.

2.3. Định nghĩa tiên đề

Các định nghĩa cổ điển và thống kê của xác suất có nhiều hạn chế để xây dựng một lý thuyết tổng quát. Khái niệm cổ điển không dùng được trong trường hợp không thể xây dựng một hệ thống đầy đủ các sự kiện đồng khả năng. Trong khi đó, tần suất chỉ là một giá trị xấp xỉ để đánh giá xác suất, chưa kể đòi hỏi là số quan sát phải rất lớn và giá trị tần suất tìm được phải lớn hơn nhiều sai số đo và cả sai số tính toán.

Chúng ta bắt đầu từ hệ thống các tiên đề dưới dạng do Kôn-mô-gô-rốp phát biểu. Các tiên đề đó (giống như các tiên đề toán học khác) được thừa nhận là đúng đắn, tất nhiên căn cứ vào kinh nghiệm cuộc sống và hoạt động thực tiễn. Cách tiếp cận này liên hệ chặt chẽ lý thuyết xác suất với lý thuyết hàm số và tập hợp. Cách xác định xác suất theo tiên đề sẽ chứa trong nó các định nghĩa cổ điển và thống kê của xác suất như là các trường hợp riêng.

Ta quay trở lại không gian các sự kiện sơ cấp Ω (xem §1), còn bản thân các phần tử là gì không quan trọng. Tiếp theo xác định hệ thống $\mathcal{A}$ các tập hợp con của Ω, các phần tử của $\mathcal{A}$ được gọi là các sự kiện ngẫu nhiên. Ta đặt cho $\mathcal{A}$ các yêu cầu hợp lý sau:

(i) $\mathcal{A}$ chứa Ω.

(ii) Nếu A và $B \in \mathcal{A}$ thì $\bar{A}$, $\bar{B}$, $A + B$, $AB \in \mathcal{A}$.

Hệ thống $\mathcal{A}$ thỏa mãn các điều kiện trên được gọi là *đại số Bun*. Nếu ta yêu cầu thêm

(iii) Nếu $A_1, A_2, \ldots, A_n, \ldots$ là các phần tử của $\mathcal{A}$, thì tổng và tích vô hạn $A_1 + A_2 + \ldots + A_n + \ldots$, $A_1A_2 \ldots A_n \ldots$ cũng thuộc $\mathcal{A}$. Nếu $\mathcal{A}$ thỏa mãn thêm điều kiện (iii) ta có một *trường Bô-ren*, hay $\sigma-$*đại số*.

Bây giờ ta đã có thể định nghĩa xác suất:

Định nghĩa. Ta gọi xác suất trên $(\Omega, \mathcal{A})$ là một hàm số xác định trên $\mathcal{A}$ có giá trị trong $[0; 1]$ và thỏa mãn 3 tiên đề

(T_1) $P(\Omega) = 1$;

(T_2) $P(A + B) = P(A) + P(B)$ (A, B xung khắc);

(T_3) Nếu dãy $\{A_n\}$ có tính chất $A_j \Rightarrow A_i$, $\forall\, i \le j$ và

$$A_1A_2\ldots A_n\ldots = V, \text{ thì } P(A_n) \xrightarrow[n\to\infty]{} 0.$$

Xuất phát từ hệ tiên đề trên có thể chứng minh được các tính chất của xác suất đã trình bày ở §1, hoặc chính chúng đã là các tính chất đó (tiên đề 1 và 2). Chú ý rằng hệ tiên đề này chưa đầy đủ: ứng với một tập Ω có thể chọn xác suất theo nhiều cách khác nhau. Người ta có thể thay tiên đề 2 và 3 bằng một tiên đề có tên là *tiên đề cộng mở rộng*:

(T_4) Nếu dãy $\{A_n\}$ có tính chất xung khắc từng đôi và $A = \sum_{n=1}^{\infty} A_n \in \mathcal{A}$ thì

$$P(A) = P(A_1) + P(A_2) + \ldots P(A_n) + \ldots = \sum_{n=1}^{\infty} P(A_n).$$

Để kết luận, có thể nói rằng cách định nghĩa xác suất ở đây nhìn từ quan điểm của lý thuyết tập hợp chính là sự đưa vào cùng với Ω một độ đo không âm, trực chuẩn, cộng tính, xác định cho mọi phần tử của tập $\mathcal{A}$. Như vậy khi định nghĩa xác suất chúng ta phải có không chỉ tập Ω các sự kiện sơ cấp ban đầu, mà còn phải có tập các sự kiện ngẫu nhiên $\mathcal{A}$ và hàm số P xác định trên đó. Tổ hợp $\{\Omega, \mathcal{A}, P\}$ sau này thường được gọi là *không gian xác suất*.

§3. XÁC SUẤT CÓ ĐIỀU KIỆN

3.1. Khái niệm

Thực ra mọi xác suất $P(A)$ đều là có điều kiện, vì sự kiện A xảy ra khi thực hiện một bộ điều kiện xác định. Tuy nhiên, nếu ngoài bộ điều kiện đó ra còn có thêm điều kiện khác thể hiện bằng việc xuất hiện B nào đó, thì người ta đưa ra một khái niệm mới: *xác suất có điều kiện của A biết rằng đã xảy ra B*, ký hiệu là $P(A \mid B)$. Bằng trực giác ta cũng thấy rằng khi có B với $P(B) > 0$ thì nói chung "khả năng" xuất hiện A cũng thay đổi; đặc biệt nếu $AB = V$ khả năng đó triệt tiêu, còn nếu $B \Rightarrow A$ thì khả năng trở thành tất yếu. Vậy là, với điều kiện đã có B, người ta xác định một cách tự nhiên khả năng xuất hiện A nào đó bằng một số tỷ lệ với $P(AB)$, tức là số có dạng $kP(AB)$, $k > 0$. Để xác định hằng số k đó, do $P(A \mid B) = kP(AB)$ là một xác suất và ta chọn $A = B$, $P(B \mid B) = 1$, nên $kP(B) = 1$. Từ đó

$$k = \frac{1}{P(B)}.$$

Định nghĩa 1. Giả sử trong một phép thử ta có $P(B) > 0$. Khi đó xác suất có điều kiện của sự kiện A nào đó, biết rằng đã có B, sẽ là một số không âm, ký hiệu là:

$$P(A|B) = \frac{P(AB)}{P(B)}. \tag{3.1}$$

Để ý rằng nói chung $P(A) \neq P(A|B)$. Ngoài ra xác suất có điều kiện có mọi tính chất của một xác suất bình thường.

Thí dụ 3.1. Gieo 2 con xúc sắc giống nhau. Tính xác suất để ta có tổng số chấm thu được bằng 6, biết rằng tổng đó là một số chẵn.

Giải. Ta đã biết $P(A) = 5/36$ (xem thí dụ 2.2, A là sự kiện xuất hiện tổng chấm bằng 6). Nếu ký hiệu B là sự kiện xuất hiện tổng chấm chẵn, thì điều kiện để tính $P(A|B)$ đã thay đổi, tổng số chẵn chỉ tương ứng với 18 kết cục của phép thử gieo 2 con xúc sắc. Từ đó $P(A|B) = 5/18$.

Thí dụ 3.2. Rút từ bộ bài tú lơ khơ 52 con lần lượt ra 2 con bài. Tìm xác suất để con thứ hai là át, biết rằng con thứ nhất đã là át.

Giải. Dễ thấy nếu ký hiệu A_i là sự kiện con thứ i là át ($i = 1,2$), thì $P(A_2|A_1) = \frac{3}{51} = \frac{1}{17}$, tương đương với việc do đã có A_1, việc tính xác suất sự kiện A_2 đưa về tính trong trường hợp chỉ còn 51 con bài với 3 con át trong đó.

Định nghĩa 2. Ta nói rằng A và B *độc lập* (thống kê), nếu

$$P(A|B) = P(A) \text{ hoặc } P(B|A) = P(B). \tag{3.2}$$

Như vậy nếu A, B độc lập việc xuất hiện sự kiện này không làm thay đổi xác suất của sự kiện kia. Tuy nhiên việc kiểm tra tính chất (3.2) trong thực tiễn rất khó khăn và trong nhiều trường hợp là không thể. Vì vậy dựa vào thực tế và trực giác mà ta thừa nhận các sự kiện độc lập trong các bài tập sau này. Công thức tương đương của (3.2), có để ý đến (3.1) là:

$$P(AB) = P(A)P(B). \tag{3.3}$$

Định nghĩa 3. Ta nói bộ sự kiện $A_1, A_2, \dots, A_n$ *độc lập* (hay *độc lập trong tổng thể*) nếu

$$P(A_{i_1}A_{i_2}\dots A_{i_k}) = P(A_{i_1})P(A_{i_2})\dots P(A_{i_k}) \tag{3.4}$$

với mọi dãy $(i_1, \dots, i_k)$ gồm các số nguyên khác nhau lấy từ $\{1, 2, \dots, n\}$.

Thí dụ 3.3. Gieo hai lần một đồng tiền và ta có 4 kết cục đồng khả năng (S – ký hiệu mặt sấp, N – mặt ngửa)

$$\Omega = \{SS, SN, NS, NN\}.$$

Rõ ràng các sự kiện $A = SS + SN$, $B = SS + NS$, $C = SS + NN$ là độc lập từng đôi do $P(A) = P(B) = P(C) = \frac{1}{2}$; còn $P(AB) = P(AC) = P(BC) = \frac{1}{4}$ thỏa mãn (3.3). Tuy nhiên chúng không độc lập trong tổng thể do

$$P(ABC) = \frac{1}{4} \neq P(A)P(B)P(C) = \frac{1}{8}.$$

Như vậy không nên nhầm lẫn hai khái niệm độc lập trong các định nghĩa 2 và 3. Khái niệm độc lập trong tổng thể kéo theo độc lập từng đôi (do (3.3) là trường hợp riêng của (3.4) khi $k = 2$), nhưng ngược lại nói chung không đúng.

3.2. Công thức cộng và nhân xác suất

1. Công thức nhân xác suất

$$P(AB) = P(A)P(B \mid A) = P(B)P(A \mid B). \tag{3.5}$$

Đó là hệ quả trực tiếp suy ra từ (3.1). Từ (3.5) có thể dẫn ra các kết quả quan trọng:

(i) Nếu A, B độc lập thì $P(AB) = P(A)P(B)$ (xem 3.3)).

(ii) Mở rộng cho tích n sự kiện

$$P(A_1A_2\ldots A_n) =$$
$$= P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\ldots P(A_n|A_1A_2\ldots A_{n-1}). \quad (3.6)$$

(iii) Nếu $A_1A_2, \ldots A_n$ độc lập trong tổng thể, thì:

$$P\left(\prod_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i).$$

2. Công thức cộng xác suất

$$P(A + B) = P(A) + P(B) - P(AB). \quad (3.7)$$

Việc chứng minh công thức trên không có gì quá phức tạp (nhất là từ các tiên đề của mục 2.3). Từ (3.7) có thể dẫn ra các kết quả sau:

(i) Nếu A, B xung khắc, thì $P(A + B) = P(A) + P(B)$.

(ii) Mở rộng cho tổng n sự kiện

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) - \ldots$$
$$+ (-1)^{n-1}P(A_1A_2\ldots A_n). \quad (3.8)$$

(iii) Nếu $A_1, A_2, \ldots A_n$ xung khắc từng đôi

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

Các công thức (3.5) – (3.8) cho ta các công cụ hiệu quả để tính xác suất các sự kiện phức tạp qua xác suất các sự kiện đơn giản hơn.

Thí dụ 3.4. Hai cọc bài được lấy từ một bộ bài tú lơ khơ, cọc thứ nhất gồm 4 con át, cọc thứ hai gồm 4 con ka. Rút ngẫu nhiên từ mỗi cọc bài ra một con bài, tính các xác suất để

a) cả 2 con là con cơ,

b) có ít nhất 1 con cơ.

Cũng câu hỏi như vậy nhưng thay điều kiện đầu bài: trộn cọc bài và rút hú họa từ đó ra 2 con bài.

Giải. Gọi A – con bài thứ nhất là cơ, B – con bài thứ hai là cơ. Để ý rằng thuật ngữ "thứ nhất"... chỉ để phân biệt hai con bài chứ không để chỉ thứ tự nào cả. Trong trường hợp hai cọc bài riêng rẽ, dễ thấy A và B độc lập. Từ đó

a) Xác suất cần tìm là $P(AB)$, để ý đến (3.3) ta có:

$$P(AB) = P(A)P(B) = \frac{1}{4}.\frac{1}{4} = \frac{1}{16}.$$

b) Sự kiện ta quan tâm là $A + B$, theo (3.7):

$$P(A+B) = P(A) + P(B) - P(AB) = \frac{1}{4} + \frac{1}{4} - \frac{1}{16} = \frac{7}{16}.$$

Trường hợp trộn lẫn hai cọc bài thành một thì A, B không còn độc lập nữa. Tuy nhiên các xác suất $P(A)$ và $P(B)$ đều bằng $2/8 = 1/4$ do vai trò hai quân bài như nhau. Từ đó:

a) Dùng công thức (3.5):

$$P(AB) = P(A)P(B|A) = \frac{1}{4}.\frac{1}{7} = \frac{1}{28}.$$

b) Một lần nữa theo (3.7):

$$P(A+B) = P(A) + P(B) - P(AB) = \frac{1}{4} + \frac{1}{4} - \frac{1}{28} = \frac{13}{28}.$$

Thí dụ 3.5. Ba xạ thủ mỗi người bắn một viên đạn với xác suất bắn trúng của từng người tương ứng là 0,7; 0,8 và 0,9. Tính các xác suất:

a) có hai người bắn trúng,

b) có ít nhất một người bắn trượt.

Giải. Gọi A_i là sự kiện xạ thủ thứ i bắn trúng (i = 1, 2, 3) và $P(A_1) = 0,7$; $P(A_2) = 0,8$; $P(A_3) = 0,9$.

a) Nếu gọi A là sự kiện có đúng 2 người bắn trúng thì:

$$A = A_1A_2\overline{A_3} + A_1\overline{A_2}A_3 + \overline{A_1}A_2A_3.$$

Dùng tính xung khắc của các số hạng và tính độc lập của các A_i và $\overline{A_j}$ $(j \neq i)$, ta có:

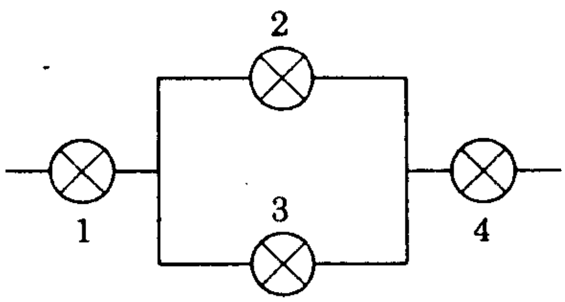
$$\begin{aligned}
P(A) &= P(A_1A_2\overline{A_3}) + P(A_1\overline{A_2}A_3) + P(\overline{A_1}A_2A_3)\\
&= P(A_1)P(A_2)P(\overline{A_3}) + P(A_1)P(\overline{A_2})P(A_3) + P(\overline{A_1})P(A_2)(A_3)\\
&= 0{,}7.0{,}8.(1 - 0{,}9) + 0{,}7.(1 - 0{,}8).0{,}9 + (1 - 0{,}7).0{,}8.0{,}9\\
&= 0{,}398.
\end{aligned}$$

b) Nếu gọi B là sự kiện có ít nhất một người bắn trượt, thì $\overline{B}$ là sự kiện không có ai bắn trượt hay cả ba đều bắn trúng. Rõ ràng việc tính $P(\overline{B})$ dễ dàng hơn nhiều so với tính $P(B)$ theo cách trực tiếp, từ đó

$$\begin{aligned}
P(B) &= 1 - P(\overline{B}) = 1 - P(A_1A_2A_3)\\
&= 1 - 0{,}7.0{,}8.0{,}9 = 0{,}496.
\end{aligned}$$

Thí dụ 3.6. Cho một mạch điện gồm 4 linh kiện như hình 1.3, trong đó xác suất hỏng của từng linh kiện trong một khoảng thời gian nào đó tương ứng là 0,2; 0,1; 0,05 và 0,02. Tìm xác suất để mạng hoạt động tốt trong khoảng thời gian đó, với giả thiết là các linh kiện làm việc độc lập với nhau và các dây luôn tốt.

Giải. Gọi A_i là sự kiện linh kiện thứ i làm việc tốt $(i = \overline{1,4})$. Sử dụng các tính chất của mạng song song và nối tiếp, gọi A là sự kiện mạng hoạt động tốt, khi đó $A = A_1(A_2 + A_3)A_4$.



Hình 1.3

Để ý rằng từ giả thiết đầu bài ta luôn có A_1, A_4 và $A_2 + A_3$ độc lập, nên:

$$P(A) = P(A_1)P(A_2 + A_3)P(A_4). \qquad (3.9)$$

Ta cần tính $P(A_2 + A_3)$, và do A_2, A_3 không xung khắc, nên

$$P(A_2 + A_3) = P(A_2) + P(A_3) - P(A_2A_3).$$

Thay vào (3.9), để ý rằng $P(A_2A_3) = P(A_2)P(A_3)$ và giả thiết của đầu bài

$$\begin{aligned} P(A) &= P(A_1)[P(A_2) + P(A_3) - P(A_2)P(A_3)]P(A_4) \\ &= 0,8.(0,9 + 0,95 - 0,9.0,95).0,98 \\ &= 0,78008. \end{aligned}$$

Chú ý rằng nếu ta khai triển $A = A_1A_2A_4 + A_1A_3A_4$ sau đó dùng các công thức (3.6) – (3.7) để tính $P(A)$ thì sẽ phức tạp hơn một chút, bạn đọc hãy tự giải theo cách này.

Thí dụ 3.7. Một gia đình có 6 con. Tìm xác suất để gia đình đó có số con trai nhiều hơn số con gái.

Giải. Ta chấp nhận xác suất sinh con trai bằng xác suất sinh con gái và bằng 0,5, ngoài ra kết quả mỗi lần sinh được coi là độc lập với nhau. Gọi A là sự kiện số con trai nhiều hơn con gái, khi đó việc tính trực tiếp $P(A)$ đưa về xác định các trường hợp: hoặc 6 trai, hoặc 5 trai 1 gái, hoặc 4 trai 2 gái. Tuy nhiên có thể dùng cách khác. Gọi B là sự kiện số gái nhiều hơn trai, còn C là sự kiện số trai và số gái như nhau. Dễ thấy

$$A + B + C = U \text{ và } P(A) + P(B) + P(C) = 1.$$

Do tính đối xứng của việc sinh con trai và con gái, nên $P(A) = P(B)$, từ đó:

$$P(A) = \frac{1 - P(C)}{2}$$

và ta cần phải tính $P(C)$ – xác suất để trong gia đình có 3 con trai, 3 con gái. Một trường hợp như vậy có xác suất $\frac{1}{2^6}$ và có tất cả $C_6^3 = 20$ khả năng khác nhau, từ đó $P(C) = 20/64 = \frac{5}{16}$ và

$$P(A) = \frac{1 - \frac{5}{16}}{2} = \frac{11}{32}.$$

Thí dụ 3.8. Một người viết n là thư cho n người khác nhau, bỏ ngẫu nhiên vào n phong bì đã có sẵn địa chỉ. Tìm xác suất để có ít nhất một lá thư bỏ vào đúng phong bì.

Giải. Gọi A_i là sự kiện là thư thứ i bỏ đúng phong bì ($i = \overline{1, n}$), A – là sự kiện cần tìm xác suất, ta có $A = A_1 + A_2 + \ldots + A_n$. Do các A_i không xung khắc, nên ta dùng công thức (3.8). Dễ thấy

$$P(A_i) = \frac{1}{n} = \frac{(n-1)!}{n!};$$

$$P(A_iA_j) = P(A_i)P(A_j|A_i) = \frac{1}{n}.\frac{1}{n-1} = \frac{(n-2)!}{n!};$$

$$P(A_iA_jA_k) = P(A_i)P(A_j|A_i)P(A_k|A_iA_j) = \frac{(n-3)!}{n!};$$

....

$$P(A_1A_2...A_n) = P(A_1)P(A_2|A_1)...P(A_n|A_1A_2...A_{n-1}) = \frac{1}{n!}.$$

Từ đó theo (3.8)

$$P(A) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) - \ldots$$

$$+ (-1)^{n-1}P(A_1A_2\ldots A_n)$$

$$= C_n^1 \frac{(n-1)!}{n!} - C_n^2 \frac{(n-2)!}{n!} + C_n^3 \frac{(n-3)!}{n!} - \ldots + (-1)^{n-1}\frac{1}{n!}$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n-1}\frac{1}{n!}.$$

Khi n khá lớn xác suất cần tìm $\approx 1 - \frac{1}{e}$.

Thí dụ 3.9. Tìm xác suất để xuất hiện ít nhất 1 lần 2 mặt chấm khi gieo n lần 2 con xúc sắc.

Giải. Xác suất để trong 1 lần gieo 2 con xúc sắc ta có hai mặt 6 chấm sẽ là $\frac{1}{36}$, và không có hai mặt 6 chấm sẽ là $1-\frac{1}{36}$. Nếu đặt A là sự kiện cần tìm, rõ ràng $\bar{A}$ là sự kiện gieo n lần 2 con xúc sắc mà không lần nào có 2 mặt 6 chấm. Từ đó

$$P(\bar{A}) = \left(1-\frac{1}{36}\right)^n \text{ và } P(A) = 1-\left(\frac{35}{36}\right)^n.$$

3.3. Công thức Béc-nu-li

Xét một dãy n phép thử độc lập giống nhau, trong mỗi phép thử chỉ có hai kết cục hoặc xảy ra A hoặc không và $P(A) = p$, $P(\bar{A}) = 1 - p = q$ không phụ thuộc vào số thứ tự của phép thử. Những bài toán thỏa mãn các yêu cầu trên được gọi là tuân theo *lược đồ Béc-nu-li* và hay gặp trong nhiều lĩnh vực ứng dụng.

Ta quan tâm đến xác suất để trong dãy n phép thử độc lập nói trên sự kiện A xuất hiện đúng k lần, ký hiệu là $P_n(k)$. Gọi B là sự kiện "trong dãy n phép thử Béc-nu-li sự kiện A xuất hiện đúng k lần", ta thấy B có thể xảy ra theo nhiều phương án khác nhau, miễn sao trong dãy các kết cục của n phép thử sự kiện A có mặt đúng k lần. Rõ ràng B sẽ là tổng của C_n^k các phương án như vậy. Còn xác suất để xảy ra một phương án, do trong dãy n phép thử độc lập sự kiện A xuất hiện đúng k lần, $\bar{A}$ xuất hiện $n-k$ lần, nên sẽ bằng p^kq^{n-k}. Từ đó ta có *công thức Béc-nu-li*

$$P(B) = P_n(k) = C_n^k p^k q^{n-k},\ k = 0, 1, \dots n. \tag{3.10}$$

Việc sử dụng công thức (3.10) sẽ đơn giản hơn nhiều việc dùng các công thức (3.5) – (3.8) và vì vậy nó có ý nghĩa thực tiễn rất lớn.

Thí dụ 3.10. Một thiết bị có 10 chi tiết đối với độ tin cậy (xác suất làm việc tốt trong một khoảng thời gian nào đó) của mỗi chi tiết là 0,9. Tìm xác suất để trong khoảng thời gian ấy có đúng 2 chi tiết làm việc tốt.

Giải. Rõ ràng ta có lược đồ Béc-nu-li, với $n = 10$, $p = 0{,}9$ và $k = 2$, áp dụng (3.10) ta có xác suất cần tìm là:

$$P_{10}(2) = C_{10}^{2}.(0{,}9)^2.(0{,}1)^8 = 3645.10^{-10}.$$

Thí dụ 3.11. Một bác sỹ có xác suất chữa khỏi bệnh là 0,8. Có người nói rằng cứ 10 người đến chữa thì có chắc chắn 8 người khỏi bệnh; điều đó có đúng không?

Giải. Câu khẳng định là sai. Ở đây có thể coi việc chữa bệnh cho 10 người là dãy 10 phép thử, trong đó A là sự kiện được chữa khỏi bệnh có $P(A) = 0{,}8$. Từ đó xác suất để trong 10 bệnh nhân đến chữa có 8 người khỏi là:

$$P_{10}(8) = C_{10}^{8}.0{,}8^8.0{,}2^2 \approx 0{,}3108.$$

Thí dụ 3.12. Tỷ lệ phế phẩm của một lô hàng là 1%. Hỏi cỡ mẫu cần chọn ra là bao nhiêu (có hoàn lại) sao cho trong mẫu có ít nhất 1 phế phẩm với xác suất lớn hơn 0,95?

Giải. Giả sử mẫu chọn ra có kích cỡ là n và việc chọn ra một sản phẩm có hoàn lại là một phép thử Béc-nu-li với $p = 0{,}01$. Rõ ràng xác suất để trong mẫu có ít nhất 1 phế phẩm sẽ là:

$$1 - (1 - p)^n = 1 - 0{,}99^n.$$

Theo yêu cầu của đầu bài

$$1 - 0{,}99^n > 0{,}95 \Leftrightarrow 0{,}05 > 0{,}99^n$$

$$\Rightarrow n > \frac{\log 0{,}05}{\log 0{,}99} \approx 296.$$

Nhiều khi ta muốn tìm xác suất để trong dãy n phép thử Béc-nu-ni sự kiện A xuất hiện với số lần từ k_1 đến k_2; dễ thấy xác suất cần tìm, ký hiệu là $P_n(k_1, k_2)$, sẽ là:

$$P_n(k_1; k_2) = \sum_{k=k_1}^{k_2} P_n(k) = \sum_{k=k_1}^{k_2} C_n^k p^k q^{n-k}. \qquad (3.11)$$

Ta có nhận xét rằng khi n và k khá lớn, việc tính toán xác suất theo (3.10) và (3.11) rất cồng kềnh và khó khăn; vì vậy người ta tìm cách tính gần đúng các xác suất đó. Có thể sử dụng các cách xấp xỉ sau đây:

(i) Nếu n rất lớn, trong khi p rất nhỏ, xác suất theo công thức (3.10) có thể xấp xỉ bằng (*xấp xỉ Poa-xông*)

$$P_n(k) \approx \frac{(np)^k}{k!} e^{-np}. \tag{3.12}$$

(ii) Nếu n lớn, nhưng p không quá bé và quá lớn, ta có *xấp xỉ chuẩn* (định lý giới hạn địa phương Moa-vrơ – Láp-la-xơ)

$$P_n(k) \approx \frac{\varphi(x_k)}{\sqrt{npq}},\ x_k = \frac{k-np}{\sqrt{npq}}, \tag{3.13}$$

trong đó $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ là hàm Gao-xơ (xem bảng 1).

(iii) Nếu n lớn, nhưng p không quá bé hoặc quá lớn thì xác suất trong (3.11) có thể xấp xỉ bằng (định lý giới hạn tích phân Moa-vrơ – Láp-la-xơ)

$$P_n(k_1;\ k_2) \approx \phi(x_2) - \phi(x_1),\ x_j = \frac{k_j - np}{\sqrt{npq}},\ j = 1, 2, \tag{3.14}$$

và trong đó $\phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ là hàm Láp-la-xơ (xem bảng 2).

Thí dụ 3.13. Xác suất sản xuất ra phế phẩm của một máy là 0,005. Tìm xác suất để trong 800 sản phẩm của máy đó có đúng 3 phế phẩm.

Giải. Rõ ràng có thể dùng xấp xỉ Poa-xông theo (3.12), với $np = 4$

$$P_{800}(3) \approx \frac{e^{-4}.4^3}{3!} = 0,1954.$$

Thí dụ 3.14. Xác suất ném trúng rổ của một cầu thủ là 0,8. Tìm xác suất để trong 100 lần cầu thủ đó:

a) ném trúng 75 lần;

b) ném trúng không ít hơn 75 lần.

Giải. Việc tính theo công thức (3.10) hoặc (3.11) của lược đồ Béc-nu-li sẽ khá phức tạp. Ta sẽ tính xấp xỉ theo (3.13) và (3.14):

$$\text{a) } P_{100}(75) \approx \frac{\varphi\left(\dfrac{75-0,8.100}{\sqrt{100.0,8.0,2}}\right)}{\sqrt{100.0,8.0,2}} = \frac{\varphi(-1,25)}{4} = 0,04565.$$

b) $P_{100}(75;\ 100) \approx \phi(5) + \phi(1,25) = 0,8943.$

§4. CÔNG THỨC BAY-ÉT

4.1. Khái niệm nhóm đầy đủ

Định nghĩa. Nhóm các sự kiện $A_1, A_2, \ldots, A_n$ ($n \geq 2$) của một phép thử được gọi là (hay tạo thành) một *nhóm đầy đủ*, nếu

(i) $A_iA_j = V$, $\forall i \neq j$ (xung khắc từng đôi),

(ii) $A_1 + A_2 + \ldots + A_n = U$.

Theo định nghĩa này ở phép thử đang xét chỉ có thể xuất hiện một sự kiện trong số n sự kiện $A_1, \ldots, A_n$ (và phải có một sự kiện). Nhóm $A_1, \ldots, A_n$ có các tính chất trên còn được gọi là một hệ thống đầy đủ.

Thí dụ 4.1. Xét phép thử gieo một con xúc sắc. Nếu ký hiệu A_i là sự kiện xuất hiện mặt i chấm ($i = \overline{1,6}$), ta có một nhóm đầy đủ $\{A_i, i = \overline{1,6}\}$. Có thể tạo thành nhiều nhóm đầy đủ khác cho phép thử này, chẳng hạn đặt $A = A_6$, từ đó $\overline{A} = A_1 + A_2 + \ldots + A_5 = \overline{A_6}$ và nhóm $\{A, \overline{A}\}$ chính là một nhóm đầy đủ.

Như vậy dễ thấy tập hợp tất cả các sự kiện sơ cấp tạo nên một nhóm đầy đủ. Tổng quát hơn tập các sự kiện tạo nên một phân hoạch của không gian Ω các sự kiện sơ cấp cũng là một nhóm đầy đủ. Tập $\{A, \overline{A}\}$, với A là sự kiện tùy ý là nhóm đầy đủ bé nhất (chỉ có 2 phần tử). Để ý $\{U, V\}$ cũng tạo nên một nhóm đầy đủ và được gọi là nhóm đầy đủ *tầm thường*.

4.2. Công thức xác suất đầy đủ

Giả sử ta có một nhóm đầy đủ các sự kiện $A_1, A_2, \ldots A_n$ và đồng thời xét một sự kiện H nào đó. Nếu đã biết các $P(A_i)$ và $P(H|A_i)$, ta có thể tính được $P(H)$. Rõ ràng từ giả thiết về nhóm đầy đủ:

$$H = A_1H + A_2H + \ldots + A_nH = \sum_{i=1}^{n} A_iH.$$

Từ đó $P(H) = P\left(\sum_{i=1}^{n} A_iH\right) = \sum_{i=1}^{n} P(A_iH)$ (do xung khắc),

và áp dụng công thức nhân (3.5):

$$P(H) = \sum_{i=1}^{n} P(A_i)P(H|A_i). \tag{4.1}$$

Công thức (4.1) có tên gọi là *công thức xác suất đầy đủ* (hay *xác suất toàn phần*).

Thí dụ 4.2. Một phân xưởng có 3 máy sản xuất cùng loại sản phẩm với tỷ lệ phế phẩm tương ứng 1%; 0,5% và 0,2%. Biết rằng máy I sản xuất ra 35%, máy II – 45% và máy III – 20% sản phẩm. Chọn hú họa ra một sản phẩm, tìm xác suất đó là phế phẩm.

Giải. Đặt M_1, M_2 và M_3 tương ứng là sự kiện sản phẩm chọn ra do máy I, II và III sản xuất. Dễ thấy $\{M_i, i = \overline{1,3}\}$ tạo nên một nhóm đầy đủ và $P(M_1) = 0{,}35$; $P(M_2) = 0{,}45$; $P(M_3) = 0{,}20$. Gọi H sự kiện rút được phế phẩm, áp dụng (4.1) để ý rằng $P(H|M_1) = 1\%$; $P(H|M_2) = 0{,}5\%$; $P(H|M_3) = 0{,}2\%$, ta có

$$P(H) = \sum_{i=1}^{3} P(M_i)P(H|M_i) =$$

$$= 0{,}35.1\% + 0{,}45.0{,}5\% + 0{,}20.0{,}2\% = 0{,}615\%.$$

Ý nghĩa của xác suất này là tỷ lệ phế phẩm của phân xưởng.

Thí dụ 4.3. Có hai hộp áo, hộp I có 10 áo trong đó có 1 phế phẩm, hộp II có 8 áo trong đó có 2 phế phẩm. Lấy hú họa 1 áo từ hộp I bỏ sang hộp II, sau đó từ hộp này chọn hú họa ra 2 áo. Tìm xác suất để cả 2 áo đó đều là phế phẩm.

Giải. Ta lập nhóm đầy đủ để làm rõ thông tin về chất lượng chiếc áo mang từ hộp I sang; gọi A – áo đó là phế phẩm, $\overline{A}$ – áo tốt. Đặt H là sự kiện 2 áo cuối chọn ra đều là phế phẩm. Rõ ràng $P(A) = \frac{1}{10}$; $P(\overline{A}) = \frac{9}{10}$; ta còn cần tính $P(H|A)$ và $P(H\backslash \overline{A})$. Dùng định nghĩa xác suất:

$$P(H|A) = \frac{C_3^2}{C_9^2} = \frac{3.2}{9.8} = \frac{1}{12};$$

$$P(H|\overline{A}) = \frac{1}{C_9^2} = \frac{2}{9.8} = \frac{1}{36}.$$

Từ đó dùng (4.1)

$$P(H) = P(A)P(H|A) + P(\overline{A})P(H|\overline{A}) = \frac{1}{10}.\frac{1}{12} + \frac{9}{10}.\frac{1}{36} = \frac{1}{30}.$$

4.3. Công thức Bay-ét

Giả sử ta có một nhóm đầy đủ $A_1, A_2, \ldots A_n$, sau đó có thêm sự kiện H nào đó. Đôi khi ta muốn xác định xác suất $P(A_i|H)$, i là một số nào đó trong $\{1, 2, \ldots, n\}$. Theo công thức nhân (3.5) ta có

$$P(A_iH) = P(A_i)P(H|A_i) = P(H)P(A_i|H).$$