1.1 Công thức cộng xác suất

Trường hợp xung khắc

1. Nếu $A_1, A_2, \ldots, A_n$ đôi một xung khắc thì ta có công thức

$$\mathbb{P}(A_1 + A_2 + \ldots + A_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \ldots + \mathbb{P}(A_n).$$

Trường hợp riêng $n = 2$: Nếu A, B xung khắc thì

$$\mathbb{P}(A + B) = \mathbb{P}(A) + \mathbb{P}(B).$$

1.1 Công thức cộng xác suất

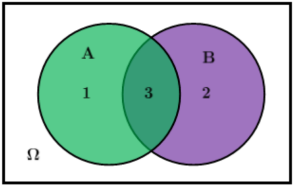
Trường hợp không xung khắc

Định lý 1.1

Cho A, B là hai biến cố tùy ý

$$\mathbb{P}(A+B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB) \tag{1.9}$$

Chứng minh:



Ta phân tích A thành tổng hai biến cố xung khắc như sau

$$A = A\Omega = A(B + \overline{B}) = AB + A\overline{B}$$

và nhận được $\mathbb{P}(A) = \mathbb{P}(AB) + \mathbb{P}(A\overline{B})$.

1.1 Công thức cộng xác suất

Trường hợp không xung khắc

Tương tự $\mathbb{P}(B) = P(AB) + \mathbb{P}(\overline{A}B)$.
Tiếp theo ta có

$$A + B = (AB + A\overline{B}) + (AB + \overline{A}B) = AB + A\overline{B} + \overline{A}B.$$

Từ đó ta thu được
$\mathbb{P}(A + B) = \mathbb{P}(AB) + \mathbb{P}(A\overline{B}) + \mathbb{P}(\overline{A}B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$.

1.1 Công thức cộng xác suất

Trường hợp không xung khắc

Hệ quả 1.3

Cho ba biến cố A, B, C. Ta có công thức

$$\mathbb{P}(A+B+C) = \mathbb{P}(A)+\mathbb{P}(B)+\mathbb{P}(C)-\mathbb{P}(AB)-\mathbb{P}(BC)-\mathbb{P}(CA)+\mathbb{P}(ABC).$$

Ví dụ 1.23

Một ngân hàng sử dụng hai loại thẻ thanh toán A và B. Tỉ lệ khách hàng sử dụng loại thẻ A là 70%, loại thẻ B là 55% và cả hai loại là 30%. Chọn ngẫu nhiên một khách hàng của ngân hàng.
Tính xác suất để:
a) Người đó sử dụng thẻ của ngân hàng.
b) Người đó không sử dụng thẻ của ngân hàng.
c) Người đó chỉ sử dụng một loại thẻ của ngân hàng.
d) Người đó chỉ sử dụng loại thẻ B của ngân hàng.

1.1 Công thức cộng xác suất

Trường hợp không xung khắc

Giải. Gọi A là biến cố *"Người đó sử dụng thẻ thanh toán A"*, B là biến cố *"Người đó sử dụng thẻ thanh toán B"*.

a) C là biến cố *"Người đó sử dụng thẻ của ngân hàng"* là biến cố $A \cup B$. Ta có biểu diễn

$$C = A + B$$

Do đó, dùng công thức cộng xác suất (1.9), ta tính được

$$\mathbb{P}(A+B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB) = 0,7 + 0,55 - 0,3 = 0,95.$$

b) Biến cố *"Người đó không sử dụng thẻ của ngân hàng"* là $\overline{A \cup B}$. Nên ta có:

$$\mathbb{P}(\overline{A+B}) = 1 - \mathbb{P}(A+B) = 1 - 0,95 = 0,05.$$

c) Biến cố *"Người đó chỉ sử dụng một loại thẻ của ngân hàng"* là biến cố $A\overline{B} \cup \overline{A}B$. Do đó,

$$\mathbb{P}(A\overline{B} + \overline{A}B) = \mathbb{P}(A\overline{B}) + \mathbb{P}(\overline{A}B) = [\mathbb{P}(A) - \mathbb{P}(AB)] + [\mathbb{P}(B) - \mathbb{P}(AB)]$$
$$= (0,7 - 0,3) + (0,55 - 0,3) = 0,4 + 0,25 = 0,65.$$

d) Biến cố *"Người đó chỉ sử dụng loại thẻ B của ngân hàng"* là biến cố $\overline{A}B$. Từ câu (c), ta có

$$\mathbb{P}(\overline{A}B) = \mathbb{P}(B) - \mathbb{P}(AB) = 0,55 - 0,3 = 0,25.$$

Một khái niệm rất quan trọng của xác suất là xác suất có điều kiện. Ta tiếp cận khái niệm này qua một ví dụ cụ thể:

Trong dịch tễ học, thay cho việc quan tâm tới tỉ lệ dân nói chung bị mắc bệnh tiểu đường, người ta quan tâm vấn đề mắc bệnh đó ở từng nhóm riêng biệt, như nhóm phụ nữ với độ tuổi từ 35 đến 50, hay nhóm những người đàn ông tuổi từ 40 đến 60.

Tương ứng ta có:

- Tỷ lệ người dân mắc tiểu đường là xác suất chọn được một người mắc tiểu đường khi chọn ngẫu nhiên một người dân.
- Tỷ lệ phụ nữ độ tuổi từ 35 đến 50 mắc tiểu đường là xác suất chọn được một người mắc tiểu đường khi chọn ngẫu nhiên một phụ nữ độ tuổi từ 35 đến 50.
- Tỷ lệ đàn ông tuổi từ 40 đến 60 mắc tiểu đường là xác suất chọn được một người mắc tiểu đường khi chọn ngẫu nhiên một đàn ông tuổi từ 40 đến 60.

1.2 Xác suất có điều kiện

Khái niệm xác suất có điều kiện

Các tỷ lệ trên là khác nhau và ta xét chúng theo cùng một phép thử là "chọn ngẫu nhiên một người dân". Tương ứng ta ký hiệu A là biến cố "người dân được chọn là một phụ nữ độ tuổi từ 35 đến 50", B là biến cố "người dân được chọn là một đàn ông độ tuổi từ 35 đến 50" , C là biến cố "người dân được chọn là một người mắc tiểu đường". Khi đó các tỷ lệ (hay là xác suất) ở trên đều được ký hiệu theo C và tương ứng là $\mathbb{P}(C), \mathbb{P}(C/A)$ và $\mathbb{P}(C/B)$.

Các xác suất $\mathbb{P}(C/A)$ và $\mathbb{P}(C/B)$ là xác suất có điều kiện.

Định nghĩa 1.9

Xác suất có điều kiện của B với điều kiện A, được ký hiệu $\mathbb{P}(B|A)$ và được xác định bởi công thức

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)}, \text{ với } \mathbb{P}(A) > 0.$$

Ví dụ 1.34

Một tập vé số có 10 vé, giả sử trong đó có 1 vé trúng thưởng. Hai người đến, lần lượt lấy mỗi người 1 vé. Tính xác suất trúng của người thứ hai trong các điều kiện người thứ nhất:

a) Lấy được vé trúng.

b) Lấy được vé không trúng.

Giải. Gọi A là biến cố chỉ *"người thứ nhất lấy được vé trúng"*, B là biến cố chỉ *"người thứ hai lấy được vé trúng"*.

a) Ta cần tính $\mathbb{P}(B|A)$. Lúc này ta hiểu là A đã xảy ra, tức người thứ nhất lấy được vé trúng. Trong tập vé số còn 9 vé và không có vé nào trúng thưởng. Khi đó cơ hội lấy được vé trúng cho người thứ hai là

$$\mathbb{P}(B|A) = \frac{0}{9} = 0.$$

b) Tương tự ta cần tính $\mathbb{P}(B|\overline{A})$. Khi biến cố đối $\overline{A}$ đã xảy ra, nghĩa là người thứ nhất lấy được vé không trúng thưởng.

Số vé còn lại gồm 9 vé, trong đó có một vé trúng thưởng. Khi đó xác suất có điều kiện của B đối với $\overline{A}$ đã xảy ra là

$$\mathbb{P}(B|\overline{A}) = \frac{1}{9}.$$

Ví dụ 1.37

Cho biết một chuyến bay thường xuyên theo lịch trình có xác suất khởi hành đúng giờ là $\mathbb{P}(D) = 0,83$ (D-Depart), xác suất đến nơi đúng giờ là $P(A) = 0,82$ (A-Arrive), và xác suất để nó vừa khởi hành và đến nơi đúng giờ là $\mathbb{P}(DA) = 0,78$. Hãy tìm xác suất để một chuyến bay:

a) Đến nơi đúng giờ, biết rằng nó khởi hành đúng giờ.

b) Khởi hành đúng giờ, biết rằng nó đến nơi đúng giờ.

Giải. a) Xác suất chuyến bay đến nơi đúng giờ, biết rằng nó khởi hành đúng giờ là $\mathbb{P}(A|D)$,

$$\mathbb{P}(A|D) = \frac{\mathbb{P}(DA)}{\mathbb{P}(D)} = \frac{0,78}{0,83} = 0,94.$$

b) Xác suất chuyến bay khởi hành đúng giờ, biết rằng nó đến nơi đúng giờ là $\mathbb{P}(D|A)$,

$$\mathbb{P}(D|A) = \frac{\mathbb{P}(DA)}{\mathbb{P}(A)} = \frac{0,78}{0,82} = 0,95.$$

Như vậy xác suất có điều kiện cho biết khả năng xảy ra của một biến cố này trong bối cảnh có thêm thông tin về biến cố khác đã xảy ra. Giả sử trong Ví dụ 1.37 cho biết thêm thông tin về một chuyến bay không khởi hành đúng giờ, khi đó khả năng nó đến nơi đúng giờ là bao nhiêu? Xác suất cần tìm là $\mathbb{P}(A|\overline{D})$.

$$\mathbb{P}(A|\overline{D}) = \frac{\mathbb{P}(A\overline{D})}{\mathbb{P}(\overline{D})} = \frac{\mathbb{P}(A) - \mathbb{P}(AD)}{1 - \mathbb{P}(D)} = \frac{0,82 - 0,78}{1 - 0,83} = 0,24.$$

Kết quả cho thấy, xác suất đến đúng giờ của một chuyến bay bị giảm khi biết chắc chắn rằng máy bay xuất phát không đúng giờ.

Trong các tính toán xác suất chúng ta thường vận dụng một tình huống đặc biệt trong quan hệ của các biến cố là tính độc lập giữa chúng. Ta làm quen với quan hệ này qua ví dụ sau:

Xét phép thử chọn ngẫu nhiên một sinh viên. Gọi A là biến cố "sinh viên được chọn có chiều cao > 170cm", B là biến cố "sinh viên được chọn có thể giao tiếp bằng tiếng Anh". Sự xuất hiện của A không gây ảnh hưởng tới khả năng xuất hiện của B, nghĩa là $\mathbb{P}(B/A) = \mathbb{P}(B)$.

Định nghĩa 1.10

Cho hai biến cố A và B. Ta nói biến cố B độc lập với biến cố A nếu

$$\mathbb{P}(B|A) = \mathbb{P}(B). \tag{1.14}$$

Nhận xét 1.2:

- Nếu có $\mathbb{P}(B|A) = \mathbb{P}(B)$ thì suy ra $\mathbb{P}(A|B) = \mathbb{P}(A)$, và ngược lại.
  Thật vậy, theo công thức xác suất có điều kiện, ta có

$$\mathbb{P}(B) \cdot \mathbb{P}(A|B) = \mathbb{P}(A) \cdot \mathbb{P}(B|A) = \mathbb{P}(AB)$$

  Nếu $\mathbb{P}(B|A) = \mathbb{P}(B)$ thì $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$ nên $\mathbb{P}(A|B) = \mathbb{P}(A)$.

- Nếu B độc lập với A thì A cũng độc lập với B và ta nói rằng "A, B độc lập với nhau".

- Nếu A và B là hai biến cố độc lập thì hai biến cố trong mỗi cặp sau cũng độc lập: A và $\overline{B}$; $\overline{A}$ và B; $\overline{A}$ và $\overline{B}$.

Ví dụ 1.38

Một thùng đựng 10 sản phẩm được bày bán, trong đó có 2 phế phẩm. Lấy lần lượt hai lần, mỗi lần một sản phẩm một cách ngẫu nhiên. Gọi A_i là biến cố *"lần thứ i lấy được sản phẩm tốt"* $(i = 1, 2)$.

Khi đó,

a) Nếu lấy có hoàn lại thì A_1 và A_2 là hai biến cố độc lập vì lần thứ nhất lấy xong lại hoàn trả lại thì thùng vẫn chứa 10 sản phẩm tốt và 2 sản phẩm phế phẩm như lúc đầu, do đó xác suất xảy ra A_1 và A_2 là như nhau:

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \frac{8}{10}.$$

b) Nếu lấy không hoàn lại thì số lượng các loại sản phẩm trong thùng có sự khác biệt giữa hai lần lấy. Do đó biến cố A_1 có ảnh hưởng đến biến cố A_2 hay A_1 và A_2 không độc lập. Tương ứng ta có

$\mathbb{P}(A_1) = \frac{8}{\ldots}$, $\mathbb{P}(A_2|A_1) = \frac{7}{\ldots}$ và $\mathbb{P}(A_2|\overline{A_1}) = \frac{8}{\ldots}$

Hệ quả 1.4

Nếu A, B là hai biến cố ứng với một phép thử nào đấy thì

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B|A) = \mathbb{P}(B)\mathbb{P}(A|B). \tag{1.15}$$

Trường hợp riêng, nếu A, B độc lập thì

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B). \tag{1.16}$$

Ví dụ 1.39

Giả sử có một hộp cầu chì chứa 20 cầu chì, trong đó có 5 cái bị lỗi. Chọn ngẫu nhiên 2 lần, mỗi lần một cầu chì, không hoàn lại. Tính xác suất để cả 2 cầu chì lấy ra đều bị lỗi.

Giải. Gọi A là biến cố *"cầu chì lấy lần đầu bị lỗi"* và B là biến cố *"cầu chì lấy lần thứ hai bị lỗi"*. Biến cố *"cả hai cầu chì lấy ra đều bị lỗi"* là biến cố AB. Ta thấy A và B là hai biến cố không độc lập. Sử dụng công thức nhân xác suất, ta có

$$\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B|A)$$

Trong lần chọn đầu tiên, hộp có 5 cái lỗi trong tổng 20 cái, nên

$$\mathbb{P}(A) = \frac{1}{4}.$$

Biến cố $B|A$ cho biết thông tin A xảy ra, tức là lần thứ nhất đã lấy một cầu chì bị lỗi ra khỏi hộp, nên hộp còn lại 4 cầu chì lỗi, trong tổng số 19 cầu chì. Vì vậy khả năng xảy ra B là

$$\mathbb{P}(B|A) = \frac{4}{19}.$$

Nên ta có,

$$\mathbb{P}(AB) = \frac{1}{4}.\frac{4}{19} = \frac{1}{19}.$$

Định lý 1.2

Nếu trong một phép thử, các biến cố $A_1, A_2, ..., A_n$ có thể xuất hiện, thì ta có

$$\mathbb{P}(A_1A_2...A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1A_2)...\mathbb{P}(A_n|A_1A_2...A_{n-1}). \tag{1.17}$$

Nếu các biến cố $A_1, A_2, ..., A_n$ là độc lập, thì ta có

$$\mathbb{P}(A_1A_2...A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)...\mathbb{P}(A_n). \tag{1.18}$$

Ví dụ 1.41

Một lô hàng có 100 sản phẩm, trong đó có 8 phế phẩm. Rút ngẫu nhiên lần lượt 4 sản phẩm. Nếu cả 4 sản phẩm này đều tốt thì lô hàng được chấp nhận. Hãy tính xác suất để lô hàng được chấp nhận trong các tình huống sau:
a) Rút không hoàn lại.
b) Rút có hoàn lại.

Giải. Gọi H là biến cố *"lô hàng được chấp nhận"*, gọi A_i là biến cố chỉ *"sản phẩm rút ở lần thứ i là tốt"*, $(i = 1, 2, 3, 4)$.
a) Khi rút không hoàn lại thì các biến cố A_i là không độc lập vì mỗi lần rút sau có số lượng sản phẩm tốt, xấu bị thay đổi so với lần rút trước. Do đó

$$\begin{aligned}\mathbb{P}(H) &= \mathbb{P}(A_1A_2A_3A_4) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1A_2)\mathbb{P}(A_4|A_1A_2A_3)\\ &= \frac{92}{100}.\frac{91}{99}.\frac{90}{98}.\frac{89}{97} = 0,7126.\end{aligned}$$

b) Khi rút có hoàn lại thì số lượng hàng tốt, xấu trong lô không có sự thay đổi, các lần lấy sau không phụ thuộc vào lần lấy trước đó hay các biến cố A_i là độc lập và xác suất để lấy được sản phẩm tốt ở mỗi lần lấy là như nhau. Do đó

$$\mathbb{P}(H) = \mathbb{P}(A_1A_2A_3A_4) = \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3)\mathbb{P}(A_4) = \left(\frac{92}{100}\right)^4 = 0,7164.$$

1.4 Công thức xác suất đầy đủ

Hệ biến cố đầy đủ

Trong nhiều bài toán chúng ta phải tính giá trị xác suất trong những tình huống khá phức tạp. Cụ thể là ta phải phân chia không gian mẫu một cách phù hợp để thực hiện các tính toán. Đây chính là việc phân chia không gian mẫu thành một hệ biến cố đầy đủ.

Định nghĩa 1.11

Cho một hệ các biến cố $\{B_1, B_2, ..., B_n\}$ của cùng một phép thử. Hệ các biến cố này là một hệ biến cố đầy đủ, nếu nó thỏa mãn hai điều kiện dưới đây:

i) $B_1 \cup B_2 \cup ... \cup B_n = \Omega$, với Ω là tập không gian mẫu.

ii) $B_i \cap B_j = \varnothing$, $\forall i \neq j$ và $i, j = \overline{1, n}$.

Nhận xét 1.3. Với mọi biến cố A của phép thử thì ta luôn có hệ hai biến cố $\{A, \overline{A}\}$ là một hệ đầy đủ.

Định lý 1.3

Giả sử $\{B_1, B_2, ..., B_n\}$ là một hệ biến cố đầy đủ của một phép thử nào đó. Giả sử A là một biến cố bất kỳ có liên quan đến phép thử. Khi đó, xác suất của A được tính theo công thức sau đây và gọi là công thức xác suất đầy đủ.

$$\mathbb{P}(A) = \mathbb{P}(B_1)\mathbb{P}(A|B_1) + \mathbb{P}(B_2)\mathbb{P}(A|B_2) + ... + \mathbb{P}(B_n)\mathbb{P}(A|B_n). \quad (1.19)$$

Chứng minh. Ta có $B_1 + B_2 + \ldots + B_n = \Omega$ nên

$$A = A\Omega = A(B_1 + B_2 + \ldots + B_n) = AB_1 + AB_2 + \ldots + AB_n.$$

Với $i \neq j$ ta có $(AB_i)(AB_j) = AB_iB_j = A\varnothing = \varnothing$ nên sử dụng công thức cộng, công thức nhân xác suất ta có

$$\begin{aligned}
\mathbb{P}(A) &= \mathbb{P}(AB_1 + AB_2 + \ldots + AB_n) \\
&= \mathbb{P}(AB_1) + \mathbb{P}(AB_2) + \ldots + \mathbb{P}(AB_n) \\
&= \mathbb{P}(B_1)\mathbb{P}(A|B_1) + \mathbb{P}(B_2)\mathbb{P}(A|B_2) + ... + \mathbb{P}(B_n)\mathbb{P}(A|B_n).
\end{aligned}$$

Ví dụ 1.47

Theo nghiên cứu của Bộ y tế cho rằng tỷ lệ hút thuốc lá là 20%. Xác suất bác sĩ chẩn đoán người hút thuốc lá mắc bệnh ung thư $0,78$. Xác suất chẩn đoán người không hút thuốc bị ung thư là $0,06$. Chọn ngẫu nhiên một người để kiểm tra. Tính xác suất để người đó bị chẩn đoán mắc bệnh ung thư.

Giải. Gọi B là biến cố người được chọn là người hút thuốc lá. Gọi A là biến cố người được chọn bị chẩn đoán mắc bệnh ung thư. Ta thấy rằng hệ $\{B, \overline{B}\}$ là một hệ đầy đủ, với $\mathbb{P}(B) = 0,2,\ \mathbb{P}(\overline{B}) = 1 - \mathbb{P}(B) = 0,8$. Tính $\mathbb{P}(A)$ theo công thức xác suất đầy đủ

$$\mathbb{P}(A) = \mathbb{P}(B)\mathbb{P}(A|B) + \mathbb{P}(\overline{B})\mathbb{P}(A|\overline{B}) = 0,2.0,78 + 0,8.0,06 = 0,204.$$

Định lý 1.4

Nếu $\{B_1, B_2, ..., B_n\}$ là một hệ đầy đủ các biến cố của một phép thử nào đó sao cho $P(B_i) \neq 0$, với $i = 1, 2, ..., n$ và A là một biến cố bất kỳ của cùng một phép thử, sao cho $P(A) \neq 0$. Khi đó ta có công thức Bayes như sau:

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_k)\mathbb{P}(A|B_k)}{\mathbb{P}(B_1)\mathbb{P}(A|B_1) + ... + \mathbb{P}(B_n)\mathbb{P}(A|B_n)}, \text{ với } k = 1, 2, ..., n. \tag{1.20}$$

Chứng minh. Công thức này được suy từ công thức xác suất có điều kiện $\mathbb{P}(B_k|A) = \frac{\mathbb{P}(B_kA)}{\mathbb{P}(A)}$. Ta thay $\mathbb{P}(B_kA)$ trên tử số bằng công thức nhân xác suất $\mathbb{P}(B_kA) = \mathbb{P}(B_k)\mathbb{P}(A|B_k)$ và thay $\mathbb{P}(A)$ dưới mẫu số bởi công thức xác suất đầy đủ (1.19), sẽ nhận được công thức (1.20).

Ví dụ 1.49

Có ba cái hộp đựng sản phẩm, hộp thứ nhất chứa 6 chính phẩm và 2 phế phẩm, hộp thứ hai chứa 10 chính phẩm và 4 phế phẩm, hộp thứ ba chứa 15 chính phẩm và 5 phế phẩm. Lấy ngẫu nhiên một hộp, rồi từ đó lấy ngẫu nhiên một sản phẩm.
a) Tính xác suất để sản phẩm lấy ra là chính phẩm.
b) Giả sử sản phẩm lấy ra là chính phẩm, hãy tính xác suất để đó là sản phẩm của hộp thứ hai.

Giải. Gọi B_i là biến cố *"hộp được lựa chọn là hộp thứ i"*, $i = 1, 2, 3$. Khi đó, $\{B_1, B_2, B_3\}$ là một hệ đầy đủ, với

$$\mathbb{P}(B_1) = \mathbb{P}(B_2) = \mathbb{P}(B_3) = \frac{1}{3}.$$

Gọi A là biến cố *sản phẩm được lấy ra là một chính phẩm*.

a) Tính $\mathbb{P}(A)$ theo công thức xác suất đầy đủ

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(B_1)\mathbb{P}(A|B_1) + \mathbb{P}(B_2)\mathbb{P}(A|B_2) + \mathbb{P}(B_3)\mathbb{P}(A|B_3)\\ &= \frac{1}{3}.\frac{6}{8} + \frac{1}{3}.\frac{10}{14} + \frac{1}{3}.\frac{15}{20} = \frac{31}{42} = 0,7381.\end{aligned}$$

b) Giả sử lấy được chính phẩm, tức là A đã xảy ra. Tính khả năng nó là sản phẩm của hộp thứ hai, nghĩa là tính $\mathbb{P}(B_2|A)$. Dùng công thức Bayes ta thu được

$$\mathbb{P}(B_2|A) = \frac{\mathbb{P}(B_2)\mathbb{P}(A|B_2)}{\mathbb{P}(A)} = \frac{\frac{1}{3}\cdot\frac{10}{14}}{\frac{31}{42}} = \frac{10}{31} = 0,3226.$$

Nhận xét 1.4. Nếu phép thử gồm hai giai đoạn, biến cố A liên quan đến giai đoạn sau, thì các kết quả có thể của giai đoạn đầu chính là một hệ đầy đủ.

Ví dụ 1.50

Một lô hàng gồm 50 sản phẩm tốt và 5 sản phẩm xấu được vận chuyển về kho, trong quá trình vận chuyển đã có 1 sản phẩm (không rõ chất lượng) bị mất. Khi lô hàng về đến kho, chọn ngẫu nhiên 1 sản phẩm:
a) Tính xác suất để sản phẩm này là sản phẩm tốt.
b) Biết rằng sản phẩm được chọn là tốt, tính xác suất để sản phẩm bị mất là sản phẩm tốt.

Giải. Gọi B là biến cố sản phẩm bị mất là sản phẩm tốt. Khi đó $\{B, \overline{B}\}$ là một hệ đầy đủ, với $\mathbb{P}(B) = \frac{50}{55}$ và $\mathbb{P}(\overline{B}) = \frac{5}{55}$. Gọi A là biến cố sản phẩm lấy ra là sản phẩm tốt.
a) Tính $\mathbb{P}(A)$ theo công thức xác suất đầy đủ. Ta cần xác định các xác suất điều kiện $\mathbb{P}(A|B)$ và $\mathbb{P}(A|\overline{B})$.

Với $\mathbb{P}(A|B)$, ta đang giả thiết tình huống B xảy ra, tức là sản phẩm bị mất là sản phẩm tốt, trong tình huống đó ta xét khả năng A xảy ra. Vậy lúc này lô hàng gồm 49 sản phẩm tốt và 5 sản phẩm xấu, nên $\mathbb{P}(A|B) = \frac{49}{54}$. Tương tự, ta có $\mathbb{P}(A|\overline{B}) = \frac{50}{54}$. Từ đó suy ra

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(B)\mathbb{P}(A|B) + \mathbb{P}(\overline{B})\mathbb{P}(A|\overline{B}) \\ &= \frac{50}{55}\cdot\frac{49}{54} + \frac{5}{55}\cdot\frac{50}{54} = \frac{50}{55} = 0,9091.\end{aligned}$$

b) Bây giờ ta giả thiết kết quả cuối của phép thử là nhận được sản phẩm tốt, tức là A xảy ra. Trong tình huống này, ta đặt câu hỏi: Sự kiện *"sản phẩm trước đó bị mất là sản phẩm tốt"* có xác suất là bao nhiêu? Tức là ta cần tính $\mathbb{P}(B|A)$. Kiểu tính này ta gọi là dùng hậu nghiệm để tính tiền nghiệm. Dùng công thức Bayes, ta có:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)} = \frac{\mathbb{P}(B)P(A|B)}{\mathbb{P}(A)} = \frac{\frac{50}{55}\cdot\frac{49}{54}}{\frac{50}{55}} = \frac{49}{54} = 0,9074.$$

1.5 Dãy phép thử độc lập và công thức Bernoulli

Dãy phép thử độc lập

Trong một số tình huống ta cần tính toán xác suất liên quan đến việc thực hiện liên tiếp một phép thử n lần độc lập. Chẳng hạn

- Ta coi việc sử dụng một máy điều hòa là thực hiện một phép thử. Tương ứng việc một công ty lắp đặt và sử dụng một lô 25 máy điều hòa chính là thực hiện $n = 25$ phép thử.
- Ta coi việc chạy một chuyến xe buýt trên một tuyến nào đó là thực hiện một phép thử. Tương ứng, việc chạy 50 chuyến xe buýt trong một ngày trên tuyến đó chính là thực hiện $n = 50$ phép thử.
- Ta coi việc theo dõi khả năng mắc bệnh truyền nhiễm ở một người dân trong một năm là thực hiện một phép thử. Tương ứng việc cơ quan y tế dự phòng theo khả năng mắc bệnh truyền nhiễm trong một năm đối với một thành phố có 500.000 dân là thực hiện $n = 500.000$ phép thử.

Định lý 1.5

Giả sử có một dãy gồm n phép thử độc lập. Với mỗi phép thử ta đều quan tâm đến biến cố A nào đó, với $\mathbb{P}(A) = p$ (không đổi). Khi đó,
a) Xác suất để trong n lần thử, biến cố A xuất hiện đúng k lần $(0 \leqslant k \leqslant n)$:

$$P_n(k;p) = C_n^k\, p^k(1-p)^{n-k}. \tag{1.21}$$

b) Xác suất để trong n lần thử, biến cố A xuất hiện từ k_1 đến k_2 lần:

$$P_n(k_1 \leqslant k \leqslant k_2;p) = \sum_{k=k_1}^{k_2} C_n^k\, p^k(1-p)^{n-k}. \tag{1.22}$$

Chứng minh. Ta gọi A_i là biến cố chỉ *"lần thử thứ i xuất hiện A"*, $i = 1, 2, .., n$. Do dãy phép thử độc lập, nên các biến cố A_i độc lập, với $\mathbb{P}(A_i) = \mathbb{P}(A) = p, \forall i = 1, 2, ..., n$.

Công thức (1.21): Gọi B là biến cố *trong n lần thử thì biến cố A xuất hiện k lần*. Tương ứng, trong n lần thử có $(n-k)$ lần không xuất hiện A, hay $(n-k)$ lần xuất hiện $\overline{A}$. Chẳng hạn

$$\underbrace{A_1A_2...A_k}_{k \text{ lần}}\underbrace{\overline{A}_{k+1}...\overline{A}_n}_{n-k \text{ lần}}.$$

Ta tách B thành một tổng của các biến cố xung khắc

$$B = \sum B_1B_2\dots B_n \tag{1.21a}$$

trong đó ứng với mỗi số hạng của vế phải $B_1B_2\dots B_n$ ta lấy

$$B_i = \begin{cases} A_i & \text{nếu lần thử thứ } i \text{ xuất hiện } A, \\ \overline{A}_i & \text{nếu lần thử thứ } i \text{ không xuất hiện } A. \end{cases}.$$

Do các lần thử độc lập nên mỗi số hạng trong vế phải của $(1.21a)$ có xác suất là

$$\mathbb{P}(B_1B_2\dots B_n) = \mathbb{P}(B_1)\mathbb{P}(B_2)\dots\mathbb{P}(B_n) = [\mathbb{P}(A)]^k[\mathbb{P}(\overline{A})]^{n-k} = p^k(1-p)^{n-k}$$

Mỗi số hạng trong vế phải của $(1.21a)$ ứng với một cách chọn k vị trí trong dãy $B_1B_2\dots B_n$ để gán $B_i = A_i$ (các vị trí khác được gán $B_i = \overline{A}_i$). Như vậy số lượng số hạng trong vế phải $(1.21a)$ chính là số tổ hợp chập k của n, hay là C_n^k. Từ đó ta có

$$\mathbb{P}(B) = C_n^k p^k (1-p)^{n-k}.$$

Người ta thường ký hiệu công thức Bernoulli theo các chỉ số có liên quan, nên $\mathbb{P}(B)$ còn được ký hiệu là $P_n(k;p)$.

Công thức (1.22): Ta gọi C là biến cố *trong n lần thử thì biến cố A xuất hiện từ k_1 đến k_2 lần*, gọi C_k là biến cố *trong n lần thử thì biến cố A xuất hiện k lần*. Như vậy biểu diễn được $C = C_{k_1} \cup C_{k_1+1} \cup \dots \cup C_{k_2}$. Do các biến cố trong tổng là xung khắc nên

$$\mathbb{P}(C) = \mathbb{P}(C_{k_1}) + \mathbb{P}(C_{k_1+1}) + \dots + \mathbb{P}(C_{k_2}).$$

Mà mỗi xác suất $\mathbb{P}(C_k)$ lại được tính như công thức (1.21). Từ đó ta nhận được công thức (1.22).

Nhận xét 1.5 Trong bài toán Bernoulli, các thông tin quan trọng cần được phân tích theo trật tự:
- Phép thử là gì?
- Phép thử đó lặp bao nhiêu lần? (tìm n)
- Mỗi lần thử ta quan tâm đến biến cố A chỉ cái gì?
- Xác suất xảy ra A là bao nhiêu? (tìm p-không đổi)
- Cần tính xác suất để trong n lần thử thì A xuất hiện mấy lần? (tìm k)

Ví dụ 1.53

Người ta kiểm tra chất lượng một thùng hàng bằng cách lấy ngẫu nhiên 5 lần, mỗi lần 1 sản phẩm, có hoàn lại. Nếu trong 5 lần lấy, có không quá 1 lần xuất hiện phế phẩm thì thùng hàng sẽ được chấp nhận. Biết rằng thùng hàng có 150 sản phẩm, trong đó có 10 phế phẩm. Tính xác suất để thùng hàng được chấp nhận.

Giải. Mỗi lần lấy một sản phẩm từ thùng hàng là một phép thử. Vì lấy có hoàn lại nên khi thực hiện 5 lần, ta nhận được một dãy phép thử độc lập

Mỗi lần thử, ta quan tâm biến cố A chỉ thông tin *sản phẩm lấy được là phế phẩm*. Ta có

$$\mathbb{P}(A) = \frac{10}{150} = \frac{1}{15}.$$

Ta thấy $\mathbb{P}(A)$ không thay đổi ở mỗi lần thử, nên $p = \frac{1}{15}$.
Gọi B là biến cố *thùng hàng được chấp nhận*. Theo đề bài, thùng hàng được chấp nhận là khi có không quá 1 phế phẩm xuất hiện trong 5 lần kiểm tra ngẫu nhiên. Vậy số lần xuất hiện phế phẩm là $0 \leqslant k \leqslant 1$. Dùng công thức Bernoulli (1.22), ta có

$$\begin{aligned}
\mathbb{P}(B) &= \sum_{k=0}^{1} C_5^k p^k (1-p)^{5-k} \\
&= C_5^0 \left(\frac{1}{15}\right)^0 \left(1 - \frac{1}{15}\right)^5 + C_5^1 \left(\frac{1}{15}\right)^1 \left(1 - \frac{1}{15}\right)^4 \\
&= 0,7082 + 0,2529 = 0,9611.
\end{aligned}$$

Ví dụ 1.55

Trong một thùng chứa 20 sản phẩm loại A, 10 sản phẩm loại B và 15 sản phẩm loại C. Lấy ngẫu nhiên 7 lần (có hoàn lại) mỗi lần một sản phẩm. Tính xác suất để trong 7 lần lấy đó
a) Có 3 lần lấy được sản phẩm loại A.
b) Có 4 lần lấy được sản phẩm loại A và 3 lần lấy được sản phẩm loại B.
c) Có 2 lần lấy được sản phẩm loại A, 4 lần lấy được sản phẩm loại B và 1 lần lấy được sản phẩm loại C.

Giải. Do lấy sản phẩm có hoàn lại nên mỗi lần lấy thì số lượng các loại sản phẩm trong thùng là như nhau và các lần lấy là độc lập với nhau. Gọi A, B, C là các biến cố chỉ sản phẩm lấy ra trong một lần lấy là sản phẩm loại A,B, C tương ứng. Ta có

$$\mathbb{P}(A) = \frac{20}{45} = \frac{4}{9}, \mathbb{P}(B) = \frac{10}{45} = \frac{2}{9}, \mathbb{P}(C) = \frac{15}{45} = \frac{1}{3}.$$

Các xác suất này không thay đổi trong mỗi lần lấy. Coi mỗi lần lấy là một phép thử ta có dãy phép thử độc lập lặp $n = 7$ lần.

a)Gọi D là biến cố "trong 7 lần lấy có 3 lần lấy được sản phẩm loại A". Theo công thức Bernoulli, ta có

$$\mathbb{P}(D) = C_7^3 \left(\frac{4}{9}\right)^3 \left(1 - \frac{4}{9}\right)^{7-3} = 0,2927.$$

b) Gọi E là biến cố "trong 7 lần lấy sản phẩm có 4 lần lấy được sản phẩm loại A và 3 lần lấy được sản phẩm loại B". Ta có

$$\mathbb{P}(E) = C_7^4 \left(\frac{4}{9}\right)^4 C_3^3 \left(\frac{2}{9}\right)^3 = 0,015.$$

c) Gọi F là biến cố "trong 7 lần lấy có 2 lần lấy được sản phẩm loại A, 4 lần lấy được sản phẩm loại B và 1 lần lấy được sản phẩm loại C". Ta có

$$\mathbb{P}(F) = C_7^2 \left(\frac{4}{9}\right)^2 C_5^4 \left(\frac{2}{9}\right)^4 C_1^1 \left(\frac{1}{3}\right)^1 = 0,0034.$$