XÁC SUẤT THỐNG KÊ

Tôn Thất Tú

Đà Nẵng, 2019

Học phần: Xác suất thống kê

Nội dung

* **Học phần** bao gồm 45 tiết, với 6 chương như sau:
- Chương 1. Xác suất
- Chương 2. Biến ngẫu nhiên
- Chương 3: Vectơ ngẫu nhiên
- Chương 4: Thống kê mô tả
- Chương 5: Ước lượng tham số
- Chương 6: Kiểm định giả thuyết thống kê

* **Tài liệu tham khảo**:

1. Jay L. Devore (2012), Probability and Statistics for Engineering and the Sciences, 8th Edition, Brooks/Cole, Cengage Learning.
2. Lê Văn Dũng (2016), Giáo trình Xác suất thống kê, NXB Thông tin và truyền thông.
3. Link: https://sites.google.com/site/tuspdn/lthuyet

* **Phần mềm**: Excel, R, Geogebra.

Chương 1: Xác suất

1. Kiến thức về tổ hợp

- Số hoán vị của một tập n phần tử:

$$P_n = n!$$

- Số cách chọn k phần tử (không thứ tự) trong tập n phần tử:

$$C_n^k = \frac{n!}{k!(n-k)!}$$

- Số cách chọn k phần tử (có thứ tự) trong tập n phần tử:

$$A_n^k = \frac{n!}{(n-k)!}$$

- **Quy tắc cộng:** Công việc A có m phương án thực hiện.
* Phương án 1: có n_1 cách
* Phương án 2: có n_2 cách
....
* Phương án m: có n_m cách
Số cách thực hiện công việc A:

$$n_1 + n_2 + ... + n_m$$

- **Quy tắc nhân:** Công việc A được thực hiện qua m giai đoạn liên tiếp.
* Giai đoạn 1: có n_1 cách
* Giai đoạn 2: có n_2 cách
....
* Giai đoạn m: có n_m cách
Số cách thực hiện công việc A:

$$n_1 * n_2 * ... * n_m$$

Ví dụ 1

Một nhóm có 10 học sinh, trong đó có 6 nam và 4 nữ. Hỏi có bao nhiêu cách:
a) Xếp thành 1 hàng dọc.
b) Chọn một nhóm có 4 học sinh đều có nam và nữ với số lượng khác nhau.
c) Chọn một nhóm có 4 học sinh trong đó có ít nhất một nữ.

Đáp số: a) 10! b) $C_6^1 C_4^3 + C_6^3 C_4^1$ c) $C_{10}^4 - C_6^4$

Ví dụ 2

Một biển số xe được cấu tạo từ 5 chữ số. Hỏi có bao nhiêu biển số có:
a) Các chữ số khác nhau.
b) Ít nhất 2 chữ số giống nhau.
c) Có đúng 3 chữ số giống nhau.

Đáp số: a) A_{10}^5 b) $10^5 - A_{10}^5$
c) Được tính qua 3 giai đoạn:
- Gđ 1: Chọn 3 vị trí trong 5 vị trí: C_5^3
- Gđ 2: Chọn 1 chữ số giống nhau cho 3 vị trí trên: C_{10}^1
- Gđ 3: chọn 2 chữ số còn lại (không phân biệt): 9^2

Vậy, theo quy tắc nhân: $C_5^3 C_{10}^1 9^2$ cách.

2. Không gian mẫu và biến cố

a. Định nghĩa

- *Thí nghiệm ngẫu nhiên*: thí nghiệm mà ta có thể lặp lại nhiều lần trong cùng điều kiện nhưng kết quả không thể dự đoán trước.
- *Không gian mẫu* của thí nghiệm: Tập tất cả các kết quả, kí hiệu Ω.
- *Biến cố* : một tập con bất kì của không gian mẫu. Kí hiệu: A,B,C, ...

Về mặt thực tế, biến cố - các hiện tượng có thể xảy ra hoặc không xảy ra trong thí nghiệm ngẫu nhiên.
- *Biến cố sơ cấp*: biến cố chỉ gồm 1 phần tử.
- Biến cố A được xem là *xảy ra* nếu có ít nhất một kết quả trong A xuất hiện.

Ví dụ 3

- *Gieo đồng xu với 2 mặt sấp, ngửa:* $\Omega = \{S, N\}$.
- *Chọn ngẫu nhiên 1 chữ số từ 0 đến 9:* $\Omega = \{0, 1, 2, ..., 8, 9\}$.
- *Gieo ngẫu nhiên 2 con xúc xắc:* $\Omega = \{(i, j) : i, j = \overline{1, 6}\}$

Ví dụ 4

Có 3 học sinh A, B, C được xếp ngẫu nhiên thành 1 hàng dọc.
Khi đó: $\Omega = \{ABC, ACB, BAC, BCA, CAB, CBA\}$. *Lúc đó, ta có biến cố:*
- A đứng giữa: $\{BAC, CAB\}$
- A không đứng giữa: $\{ABC, ACB, BCA, CBA\}$

Ví dụ 5

Một xạ thủ bắn ngẫu nhiên vào 1 tấm bia cho đến khi trúng đích.
Kí hiệu: 1 - trúng, 0 -trật. Khi đó:

$$\Omega = \{1, 01, 001, 0001, 00001, ...\}$$

Lúc đó, ta có biến cố:
- bắn trúng đích không quá 3 lần: $A = \{1, 01, 001\}$
- thực hiện ít nhất 2 lần bắn: $B = \{01, 001, 0001, ...\}$

Các biến cố đặc biệt:
- *Biến cố không thể* ($\emptyset$): là biến cố không thể xảy ra
- *Biến cố chắc chắn* (Ω): là biến cố luôn xảy ra

Chẳng hạn, gieo ngẫu nhiên 2 con xúc xắc. Lúc đó:
- biến cố "tổng số chấm xuất hiện lớn hơn 12" là biến cố không thể.
- biến cố "tổng số chấm xuất hiện lớn hơn 1" là biến cố chắc chắn.

b. Các phép toán trên biến cố

Cho hai biến cố A và B. Khi đó:
- *Biến cố đối* của A, kí hiệu $\bar{A}$ hoặc A^c, xác định: $\bar{A} = \Omega \backslash A$ (xảy ra nếu A không xảy ra)
- *Giao* của A và B, kí hiệu $A \cap B, AB$, xác định $AB = \{x : x \in A, x \in B\}$ (xảy ra nếu A và B đồng thời xảy ra).
- *Hợp* của A và B, kí hiệu $A \cup B$, xác định $A \cup B = \{x : x \in A$ hoặc $x \in B\}$ (xảy ra nếu ít nhất A hoặc B xảy ra)
- Hai biến cố A và B được gọi là *xung khắc* nếu chúng không đồng thời xảy ra.

Nhận xét
- Hai biến cố A và B xung khắc khi và chỉ khi $AB = \emptyset$.
- Hai biến cố đối nhau thì xung khắc với nhau, nhưng điều ngược lại nói chung là không đúng.

Ví dụ 6

Chọn ngẫu nhiên một chữ số từ 0 đến 9.
Ta có $\Omega = \{0, 1, 2, ..., 9\}$.
Xét 3 biến cố $A = \{0, 2, 4\}$, $B = \{4, 5, 6\}$ *và* $C = \{7, 8, 9\}$.
Lúc đó:
- $A \cap B = \{4\}, A \cup B = \{0, 2, 4, 5, 6\}, \bar{A} = \{1, 3, 5, 6, 7, 8, 9\}$
- *A và C xung khắc vì* $A \cap C = \emptyset$ *nhưng A và C không phải là hai biến cố đối nhau.*

Ví dụ 7

Có 2 xạ thủ, mỗi người bắn 1 viên đạn vào mục tiêu. Gọi A và B tương ứng là các biến cố: "người thứ nhất và thứ hai bắn trúng mục tiêu" tương ứng. Khi đó ta có biểu diễn các biến cố như sau:
- Có đạn trúng đích: $A \cup B$
- Có đúng 1 viên đạn trúng đích: $A\bar{B} \cup \bar{A}B$
- Chỉ có người thứ nhất bắn trúng: $A\bar{B}$
- Có nhiều nhất một viên đạn trúng đích: $\bar{A}\bar{B} \cup A\bar{B} \cup \bar{A}B$ *hoặc* $\overline{AB}$

Luật De-Morgan về phủ định:

$$\overline{\cup A_i} = \cap \bar{A}_i$$

$$\overline{\cap A_i} = \cup \bar{A}_i$$

3. Định nghĩa và các tính chất của xác suất

Định nghĩa (theo hệ tiên đề):
Cho trước một thí nghiệm với không gian mẫu Ω. Khi đó, **xác suất** của một biến cố A, kí hiệu $P(A)$, là số đo khả năng xảy ra biến cố A.
Ứng với mỗi biến cố A ta đặt tương ứng với giá trị $P(A)$ thỏa các điều kiện sau:
(i) Với mọi biến cố $A, P(A) \geq 0$
(ii) $P(\Omega) = 1$
(iii) Nếu $A_1, A_2, ...$ là các biến cố đôi một xung khắc thì

$$P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{+\infty} P(A_n)$$

Lúc đó $P(A)$ được gọi là *xác suất* của biến cố A.
Nhận xét: Trong một số trường hợp, tùy vào không gian mẫu cũng như các thiết lập "tương ứng" P mà ta sẽ thu được một vài định nghĩa xác suất khác.

Định nghĩa (quan điểm thống kê):
Xét biến cố A. Thực hiện phép thử n lần thì có m lần xuất hiện biến cố A. Khi đó tỉ số $f_n = m/n$ được gọi là *tần suất* xuất hiện A. Khi số phép thử tăng lên vô hạn, tần suất f_n sẽ tiến đến một hằng số xác định. Hằng số này được gọi là *xác suất* của biến cố A. Vì thế, trong thực tế khi số phép thử n lớn, ta có thể xem tần suất f_n như là xác suất của hiện tượng A.

Ví dụ 8

Có 1000 người có triệu chứng A đến cơ sở y tế để khám bệnh. Kết quả có thấy 700 mắc bệnh X. Ta có $f = 700/1000 = 70\%$. *Do đó, ta có cơ sở dự đoán nếu một người có triệu chứng A thì xác suất mắc bệnh X xấp xỉ 70%.*

Định nghĩa (quan điểm cổ điển):
Cho thí nghiệm với $n(\Omega)$ kết quả đồng khả năng, trong đó có n(A) kết quả thuận lợi cho biến cố A. Khi đó, *xác suất* của biến cố A, kí hiệu P(A), được xác định:

$$P(A) = \frac{n(A)}{n(\Omega)}$$

Ví dụ 9

Trong một hộp có 6 bi đen và 4 bi trắng. Lấy ngẫu nhiên 4 viên. Tìm xác suất lấy được cả hai loại bi.

Giải.
Gọi A là biến cố "lấy được cả hai loại bi".
Số trường hợp đồng khả năng: $n(\Omega) = C_{10}^4$.
Số trường hợp thuận lợi: $n(A) = C_6^1 C_4^3 + C_6^2 C_4^2 + C_6^3 C_4^1$.
Xác suất cần tìm:

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{C_6^1 C_4^3 + C_6^2 C_4^2 + C_6^3 C_4^1}{C_{10}^4}$$

Ví dụ 10

Một đoàn tàu có 3 toa. Có 15 khách lên ngẫu nhiên 3 toa tàu. Biết mỗi toa đều chứa được 15 khách. Tính xác suất:
a. Toa 1 có 4 khách.
b. Có 2 toa mỗi toa có 6 khách.

Giải.
a. Gọi A là biến cố "toa 1 có 4 khách".
Ta có: $n(\Omega) = 3^{15}$.
Giá trị $n(A)$ được tính thông qua 2 giai đoạn:
- Gđ 1: Chọn 4 khách: C_{15}^4 cách
- Gđ 2: Xếp 11 khách còn lại: 2^{11} cách.

Suy ra: $n(A) = C_{15}^4 2^{11}$
Xác suất cần tìm:

$$P(A) = \frac{n(A)}{n(\Omega)} = \frac{C_{15}^4 2^{11}}{3^{15}}$$

b. Gọi B là biến cố "có 2 toa mỗi toa có 6 khách".
Giá trị $n(B)$ được tính thông qua các giai đoạn:
- Gđ 1: Chọn 2 toa: C_3^2 cách
- Gđ 2: Chọn 6 khách cho mỗi toa: $C_{15}^6 C_9^6$ cách.
- Gđ 3: Xếp 3 khách còn lại $1^3 = 1$ cách

Suy ra: $n(B) = C_3^2 C_{15}^6 C_9^6$
Xác suất cần tìm:

$$P(B) = \frac{n(B)}{n(\Omega)} = \frac{C_3^2 C_{15}^6 C_9^6}{3^{15}}$$

Ví dụ 11

Một rổ cam gồm 12 quả, trong đó có 3 quả hỏng. Chia đều 12 quả này cho 3 người, mỗi người 4 quả. Tính xác suất:
a. Người thứ nhất không có quả hỏng.
b. Mỗi người đều có quả hỏng.

Gợi ý. Ta có: $n(\Omega) = C_{12}^4 C_8^4 C_4^4$
a. $n(A) = C_9^4 C_8^4 C_4^4$ b. $n(B) = 3! C_9^3 C_6^3 C_3^3$

Định nghĩa (quan điểm hình học):
Cho miền Ω đo được (trên đường thẳng, trong mặt phẳng, không gian ba chiều, ...) và miền con S đo được của Ω. Ta lấy ngẫu nhiên một điểm trong miền Ω và đặt A là biến cố $M \in S$. Khi đó, xác suất của biến cố A được xác định như sau:

$$P(A) = \frac{m(S)}{m(\Omega)},$$

trong đó $m(S), m(\Omega)$ là số đo của miền S và Ω. Cụ thể,
- Nếu Ω là đường cong hay đoạn thẳng thì $m(\cdot)$ là hàm chỉ độ dài.
- Nếu Ω là hình phẳng hay mặt cong thì $m(\cdot)$ là hàm chỉ diện tích.
- Nếu Ω là hình khối 3 chiều thì $m(\cdot)$ là hàm chỉ thể tích.
- ...

Ví dụ 12

Tính xác suất khi lấy ngẫu nhiên một điểm M trong hình vuông có độ dài cạnh bằng 2m thì điểm này rơi vào hình tròn nội tiếp hình của vuông.

Giải. Dễ thấy:
- Diện tích hình tròn: $m(S) = \pi(m^2)$
- Diện tích của hình vuông: $m(\Omega) = 4(m^2)$

Theo quan điểm hình học, xác suất của biến cố A cần tìm:

$$P(A) = \frac{m(S)}{m(\Omega)} = \frac{\pi}{4}$$

Tính chất

i) $P(\emptyset) = 0, P(\Omega) = 1$

ii) $0 \le P(A) \le 1, \forall A$

iii) Nếu $A \subset B$ thì $P(A) \le P(B)$

iv) Công thức cộng:

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC)$$

$$\begin{aligned} P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + ... + P(A_n) \\ - \sum_{i<j} P(A_i A_j) \\ + \sum_{i<j<k} P(A_i A_j A_k) \\ ...... \\ (-1)^{n-1} P(A_1 A_2 ... A_n) \end{aligned}$$

Khi các biến cố **xung khắc đôi một** thì

$$P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + ... + P(A_n)$$

Nói riêng:

$$P(A) + P(\bar{A}) = 1$$

Ví dụ 13

Một lớp có 20 sinh viên, trong đó có 10 sinh viên biết tiếng Anh, 12 sinh viên biết tiếng Pháp và 7 sinh viên biết cả hai thứ tiếng. Chọn ngẫu nhiên một sinh viên. Tìm xác suất để sinh viên đó biết ít nhất một ngoại ngữ.

Giải.
Gọi A, B là biến cố chọn được sinh viên biết tiếng Anh, Pháp.
Lúc đó, xác suất cần tìm:

$$P(A \cup B) = P(A) + P(B) - P(AB) = 10/20 + 12/20 - 7/20 = 3/4$$

Ví dụ 14

Một lô hàng có 15 thiết bị, trong đó có 6 thiết do nhà máy X sản xuất và 9 thiết bị do nhà máy Y sản xuất. Người ta chọn ngẫu nhiên 4 thiết bị để kiểm tra. Tính xác suất:
a. Cả 4 thiết bị được chọn do cùng nhà máy sản xuất
b. Có ít nhất một thiết bị được chọn do nhà máy X sản xuất.

Giải. a. Gọi A_X, A_Y là các biến cố các thiết bị được chọn do nhà máy X, Y sản xuất. Ta có A_X, A_Y xung khắc.
Xác suất cả 4 thiết bị được chọn do cùng nhà máy sản xuất:

$$P(A_X \cup A_Y) = P(A_X) + P(A_Y) = \frac{C_6^4}{C_{15}^4} + \frac{C_9^4}{C_{15}^4}$$

b. Gọi B là biến cố có ít nhất một thiết bị được chọn do nhà máy X sản xuất. Ta có:

$$P(B) = 1 - P(\bar{B}) = 1 - \frac{C_9^4}{C_{15}^4}$$

Ví dụ 15

Một đội bóng bàn của 1 đơn vị gồm 2 vận động viên A và B. Xác suất A, B vượt qua vòng bảng lần lượt là 0,7 và 0,5. Do ảnh hưởng tâm lý nên xác suất cả hai người đều vượt qua vòng bảng là 0,4. Tính xác suất cả hai vận động viên đều không vượt qua vòng bảng.

Giải. Gọi A, B là các biến cố vận động viên A, B vượt qua vòng bảng. Ta có:

$$P(A) = 0,7; P(B) = 0,5; P(AB) = 0,4$$

Xác suất cả hai vận động viên đều không vượt qua vòng bảng:

$$P(\bar{A}\bar{B}) = 1 - P(A \cup B) = 1 - [P(A) + P(B) - P(AB)]$$

$$= 1 - (0,7 + 0,5 - 0,4) = 0,2$$

4. Xác suất có điều kiện

Một lô hàng tivi với các số liệu sau:

	LG	SONY
Tốt	15	12
Bị hỏng	5	8

Chọn ngẫu nhiên 1 tivi thì được tivi LG. Vậy, khả năng nó bị hỏng là bao nhiêu?
Gợi ý.
- Khi không có thông tin, dễ thấy xác suất này là 13/40.
- Vì biết chọn được tivi LG nên ta "thu hẹp" phạm vi quan sát và chỉ tính đến loại tivi LG. Lúc đó, khả năng chọn được tivi hỏng với "điều kiện" là tivi LG bằng 5/20=1/4.

BIẾT ĐIỀU KIỆN XẢY RA $\rightarrow$ XÁC SUẤT THAY ĐỔI

Định nghĩa:

Cho hai biến cố A và B với $P(B) > 0$. Khi đó, xác suất có điều kiện của A với điều kiện biến cố B đã xảy ra, kí hiệu $P(A|B)$, xác định như sau:

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

Tính chất:

i. $0 \leq P(A|B) \leq 1$

ii. $P(\bar{A}|B) = 1 - P(A|B)$

iii. $P(A \cup B|C) = P(A|C) + P(B|C) - P(AB|C)$

Ví dụ 16

Một công ty đấu thầu 2 dự án A và B. Xác suất thắng thầu dự án A và B tương ứng là 0,6 và 0,7. Xác suất thắng thầu đồng thời cả 2 dự án là 0,5. Tính xác suất:
a) Công ty thắng thầu dự án A biết đã thắng thầu dự án B.
b) Công ty không thắng thầu dự án B biết đã thắng thầu dự án A.

Giải. Gọi A, B là các biến cố công ty thắng dự án A, B. Theo giả thiết:

$$P(A) = 0,6; P(B) = 0,7; P(AB) = 0,5$$

a. $P(A|B) = P(AB)/P(B) = 5/7$
b. $P(\bar{B}|A) = 1 - P(B|A) = 1 - P(AB)/P(A) = 1 - 5/6 = 1/6$.

Bình luận. Ta có: $P(A|B) = 5/7 > P(A) = 0,6$. Như vậy, việc biết biến cố B xảy ra đã làm tăng xác suất xảy ra biến cố A, hay nói cách khác, việc thắng dự án B là có lợi cho quá trình đấu thầu dự án A.

Nhận xét

Trong một số trường hợp, xác suất có điều kiện $P(A|B)$ có thể được tính dựa vào việc "đếm trực quan", tức là ta thu hẹp tập các kết quả sau khi biến cố B xảy ra và đếm lại số kết quả đồng khả năng, số kết quả thuận lợi để tính xác suất có điều kiện.

Chẳng hạn, ta quay lại với ví dụ về tivi Sony và LG. Vì biết chọn được tivi LG nên tập kết quả thu hẹp về 20 tivi LG. Chỉ có 5 tivi LG bị hỏng, nên xác suất chọn được tivi hỏng khi biết đó là tivi bằng 5/20=1/4.

Ví dụ 17

Có 15 thanh RAM, trong đó có 3 thanh bị hỏng. Chọn ngẫu nhiên 4 thanh.
a. Tính xác suất có ít nhất 1 thanh bị hỏng được chọn.
b. Biết chọn được ít nhất 1 thanh hỏng, tính xác suất chọn đúng 2 thanh bị hỏng.

Giải. Gọi A, B là các biến cố chọn được ít nhất 1 thanh hỏng, chọn được đúng 2 thanh hỏng.
a. $P(A) = 1 - P(\bar{A}) = 1 - C_{12}^4/C_{15}^4$
b. $P(B|A) = P(AB)/P(A) = P(B)/P(A)$, trong đó $P(B) = C_3^2 C_{12}^2/C_{15}^4$.

5. Công thức nhân xác suất

Cho hai biến cố A và B với $P(A) > 0$. Khi đó:

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

Tương tự, ta sẽ có:

$$P(AB) = P(A)P(B|A) = P(B)P(A|B)$$

Tổng quát: Cho n biến cố A_i với $P(A_1A_2...A_{n-1}) > 0$. Khi đó:

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

Ví dụ 18

Một công ty đấu thầu 2 dự án A và B. Xác suất thắng thầu lần lượt là 0.7 và 0.4. Nếu dự án A đã thắng thầu thì xác suất thắng tiếp dự án B là 0.4. Tính xác suất:
a. Công ty thắng ít nhất một dự án.
b. Công ty chỉ thắng dự án A.
c. Thắng đúng 1 dự án.

Giải. Gọi A, B là biến cố công ty thắng dự án A, B. Ta có:

$$P(A) = 0,7; P(B) = 0,4, P(B|A) = 0,4$$

a. $P(A \cup B) = P(A) + P(B) - P(AB) = P(A) + P(B) - P(A)P(B|A) = 0,7 + 0,4 - 0,7 * 0,4 = 0,82$
b. $P(A\bar{B}) = P(A)P(\bar{B}|A) = P(A)[1 - P(B|A)] = 0,7[1 - 0,4] = 0,42$
c. $P(A\bar{B} \cup \bar{A}B) = P(A\bar{B}) + P(\bar{A}B) = \ldots$

Ví dụ 19

Một thủ kho có một chùm chìa khóa gồm 9 chiếc, bề ngoài chúng giống hệt nhau nhưng trong đó chỉ có đúng 2 chiếc mở được kho. Anh ta thử ngẫu nhiên từng chìa cho đến khi mở được kho thì dừng. Tính xác suất việc làm này:
a. Dừng ở lần thử thứ 2.
b. Dừng ở lần thử thứ 3.
c. Sau nhiều nhất hai lần thử.

Giải. Gọi A_i là biến cố lần thứ i chọn đúng chìa khóa, $i = 1, 2, 3, ...$
a. $P(\bar{A}_1 A_2) = P(\bar{A}_1)P(A_2|\bar{A}_1) = 7/9 * 2/8$
b. $P(\bar{A}_1\bar{A}_2A_3) = P(\bar{A}_1)P(\bar{A}_2|\bar{A}_1)P(A_3|\bar{A}_1\bar{A}_2) = 7/9 * 6/8 * 2/7$
c. $P(A_1 \cup \bar{A}_1A_2) = P(A_1) + P(\bar{A}_1A_2) = 2/9 + 7/9 * 2/8$

6. Biến cố độc lập

- Hai biến cố A và B được gọi là độc lập nếu $P(AB) = P(A).P(B)$.
- Dãy n biến cố $A_1, ..., A_n$ được gọi là độc lập nếu ta lấy ra một dãy bất kì các biến cố khác nhau từ các biến cố trên thì xác suất của tích các biến cố này bằng tích xác suất của từng biến cố thành phần, nghĩa là:

$$P(\bigcap_{k\in I} A_k) = \prod_{k\in I} P(A_k), \forall I \subset \{1, 2, ..., n\}$$

Nhận xét

- Hai biến cố A và B độc lập $\Leftrightarrow P(A|B) = P(A)$ hoặc $P(B|A) = P(B)$.
- Nếu nhóm n biến cố $A_1, ..., A_n$ độc lập thì nhóm n biến cố $B_1, B_2, ..., B_n$ cũng độc lập, trong đó B_i là A_i hoặc $\bar{A}_i$.
- Nếu nhóm n biến cố $A_1, ..., A_n$ độc lập thì mọi nhóm con các biến cố của nó cũng độc lập.

Ví dụ 20

Một nồi hơi có 3 van bảo hiểm hoạt động độc lập với xác suất hỏng của van 1, van 2, van 3 trong khoảng thời gian T tương ứng là 0,1; 0,2; 0,3. Nồi hơi hoạt động an toàn nếu có ít nhất một van không hỏng. Tính xác suất để nồi hơi hoạt động an toàn trong khoảng thời gian T.

Giải. Gọi A_i là biến cố van i bị hỏng trong khoảng thời gian T, $i = 1, 2, 3$.
Ta có A_1, A_2, A_3 độc lập và

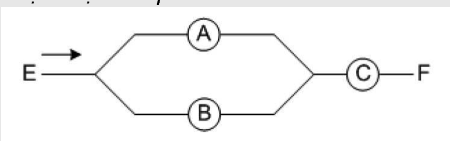
$$P(A_1) = 0,1; P(A_2) = 0,2; P(A_3) = 0,3$$

Xác suất để nồi hơi hoạt động an toàn trong khoảng thời gian T:

$$P(\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3) = 1 - P(A_1A_2A_3) = 1 - P(A_1)P(A_2)P(A_3) = 0,994$$

Ví dụ 21

Một trạm cấp nước như hình vẽ:



Nước được cấp từ E đến F qua 3 trạm bơm tăng áp A, B, C. Các trạm bơm làm việc độc lập với nhau. Xác suất để các trạm này có sự cố sau một thời gian làm việc lần lượt là 0.1, 0.1, 0.05. Tính xác suất để vùng F bị mất nước.

Giải. Gọi A,B,C là các biến cố trạm bơm A,B,C bị sự cố.
Ta có: A,B,C là các biến cố độc lập và

$$P(A) = 0,1; P(B) = 0,1; P(C) = 0,05$$

Xác suất vùng F bị mất nước:

$$P(AB \cup C) = P(AB) + P(C) - P(ABC) = P(A)P(B) + P(C) - P(A)P(B)P(C)$$

Ví dụ 22

Một thiết bị có 2 bộ phận hoạt động độc lập. Xác suất bộ phận thứ nhất bị hỏng là 0,1. Xác suất có đúng 1 bộ phận bị hỏng là 0,26.
a. Tính xác suất bộ phận thứ 2 bị hỏng.
b. Biết có ít nhất 1 bộ phận hỏng, tính xác suất bộ phận 1 hỏng.

Giải. Gọi A_i là biến cố bộ phận thứ i bị hỏng, $i = 1, 2$. Ta có A_1, A_2 độc lập.
a. Theo giả thiết:
$P(A_1\bar{A}_2 \cup \bar{A}_1A_2) = P(A_1\bar{A}_2) + P(\bar{A}_1A_2)$
$= 0,1 * [1 - P(A_2)] + 0,9P(A_2) = 0,1 + 0,8P(A_2) = 0,26.$
Suy ra: $P(A_2) = 0,2.$
b. $P(A_1|A_1 \cup A_2) = \dfrac{P(A_1(A_1 \cup A_2))}{P(A_1 \cup A_2)} = \dfrac{P(A_1)}{P(A_1 \cup A_2)}$
$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1A_2) = 0,1 + 0,2 - 0,1 * 0,2 = 0,28$
$\Rightarrow P(A_1|A_1 \cup A_2) = \dfrac{0,1}{0,28} = \dfrac{5}{14}$

7. Công thức xác suất toàn phần và công thức Bayes

Định nghĩa: Nhóm biến cố $H_i, i = \overline{1,n}$ được gọi là đầy đủ nếu thỏa mãn 2 điều kiện:
i) $H_i \cap H_j = \emptyset, \ \forall i \neq j$ ii) $\bigcup_{i=1}^{n} H_i = \Omega$

Ví dụ 23

a. $\{A, \overline{A}\}, \{\emptyset, \Omega\}$ là các nhóm biến cố đầy đủ.
b. Một hộp có 3 bi đỏ và 2 bi xanh. Lấy ngẫu nhiên 2 viên bi. Gọi H_i là biến cố lấy được i bi xanh, $i = 0, 1, 2$. Ta có $\{H_0, H_1, H_2\}$ là nhóm đầy đủ.

Định lý: (Công thức xác suất đầy đủ)

Cho $H_i, i = \overline{1,n}$ là một nhóm biến cố đầy đủ và $P(H_i) > 0$. Khi đó với biến cố A bất kì liên quan đến phép thử, ta luôn có:

$$P(A) = \sum_{i=1}^{n} P(H_i).P(A|H_i)$$

Công thức trên được gọi là *công thức xác suất đầy đủ* (hay *công thức xác suất toàn phần*).
Nói riêng, ta có $P(B) = P(A).P(B|A) + P(\overline{A}).(B|\overline{A})$ với $0 < P(A) < 1$.

Ví dụ 24

Có 2 lô sản phẩm. Lô 1 có 50 sản phẩm trong đó có 20 sản phẩm xấu. Lô 2 có 40 sản phẩm, trong đó có 15 sản phẩm xấu. Lấy ngẫu nhiên một lô và từ đó lấy ra 1 sản phẩm. Tìm xác suất để sản phẩm lấy ra là sản phẩm tốt.

Giải. Gọi H_i là chọn lô sản phẩm $i, i = 1, 2$. Gọi A là biến cố sản phẩm lấy ra là sản phẩm tốt.
Ta có $\{H_1, H_2\}$ là nhóm đầy đủ. Theo công thức xác suất toàn phần:

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = 1/2 * 30/50 + 1/2 * 25/40 = 49/80$$

Ví dụ 25

Một nhóm có 3 người nhưng chỉ có 2 vé xem bóng đá. Để chia vé họ làm như sau: Lấy 3 phiếu, 2 phiếu ghi số 1 và 1 phiếu ghi số 0. Sau đó thay phiên nhau bốc ngẫu nhiên lần lượt không hoàn lại. Ai được phiếu ghi số 1 thì được vé.
a. Tính xác suất người thứ 2 được vé.
b. Hỏi việc bốc phiếu đó có công bằng hay không ?

Giải. Gọi A_i là biến cố người rút thứ i được vé,$i = 1, 2, 3$.
a. Ta có $\{A_1, \bar{A}_1\}$ là nhóm đầy đủ. Do đó:

$$P(A_2) = P(A_1)P(A_2|A_1) + P(\bar{A}_1)P(A_2|\bar{A}_1) = 2/3 * 1/2 + 1/3 * 2/2 = 2/3$$

b. Ta có $P(A_1) = P(A_2) = 2/3$, cần tính $P(A_3)$. Mặt khác,

$$P(A_3) = 1 - P(\bar{A}_3) = 1 - P(A_1A_2) = 1 - P(A_1)P(A_2|A_1) = 1 - 2/3 * 1/2 = 2/3$$

Vậy, việc làm trên là công bằng.

Ý nghĩa công thức xác suất toàn phần

Công thức xác suất toàn phần giúp ta tính xác suất xảy ra của một biến cố dựa vào một nhóm đầy đủ các giả thiết chi phối nó.

Định lý (Công thức Bayes)

Cho A là một biến cố và $P(A) > 0$, $\{H_1, ..., H_n\}$ là một nhóm biến cố đầy đủ. Lúc đó, ta có công thức:

$$P(H_i|A) = \frac{P(H_i).P(A|H_i)}{P(A)} = \frac{P(H_i).P(A|H_i)}{\sum\limits_{j=1}^{n} P(H_j).P(A|H_j)}, i = \overline{1,n}$$

Ví dụ 26

Một trạm chỉ phát hai tín hiệu A và B với xác suất tương ứng 0,85 và 0,15. Do có nhiễu trên đường truyền nên 1/7 tín hiệu A bị méo và thu được như tín hiệu B; còn 1/8 tín hiệu B bị méo và thu được như A.
a. Tìm xác suất thu được tín hiệu A.
b. Giả sử đã thu được tín hiệu A. Tìm xác suất thu được đúng tín hiệu lúc phát.

Giải. Gọi H_A, H_B là biến cố trạm phát tín hiệu A, B. Gọi A là biến cố trạm thu được tín hiệu A.
a. Ta có: $\{H_A, H_B\}$ là nhóm biến cố đầy đủ. Theo công thức xác suất toàn phần:

$$P(A) = P(H_A)P(A|H_A) + P(H_B)P(A|H_B) = 0,85 * (1 - 1/7) + 0,15 * 1/8 = 0,747$$

b. Theo công thức Bayes:

$$P(H_A|A) = \frac{P(H_A)P(A|H_A)}{P(A)} = \frac{0,85 * (1 - 1/7)}{0,747} = 0,975$$

Ví dụ 27

Một cửa hàng bán bóng đèn cùng loại do 3 cơ sở sản xuất cung cấp. Cơ sở I, II, III cung cấp lượng hàng tương ứng là 40%, 35%, 25%. Biết tỉ lệ bóng hỏng do cơ sở I, II, III sản xuất lần lượt là 2%, 2%, 3%. Ta mua ngẫu nhiên 1 bóng của cửa hàng. Giả sử bóng mua bị hỏng. Hỏi bóng ta mua có khả năng do cơ sở nào sản xuất nhất ?

Giải. Gọi H_i là biến cố bóng được mua do cơ sở i sản xuất, $i = 1, 2, 3$. Gọi A là biến cố bóng được mua bị hỏng.
Ta có: $\{H_1, H_2, H_3\}$ là nhóm biến cố đầy đủ. Theo công thức xác suất toàn phần:

$$\begin{aligned} P(A) &= P(H_1)P(A|H_1) + P(H_2)P(A|H_2) + P(H_3)P(A|H_3) \\ &= 0,4 * 0,02 + 0,35 * 0,02 + 0,25 * 0,03 = 0.0225 \end{aligned}$$

Theo công thức Bayes: $P(H_1|A) = \dfrac{P(H_1)P(A|H_1)}{P(A)} = \dfrac{0,4 * 0,02}{0.0225} = 0.356$
Tương tự: $P(H_2|A) = 0.311, P(H_3|A) = 0.333$
Vì $P(H_1|A) > P(H_3|A) > P(H_2|A)$ nên khả năng bóng được mua do cơ sở 1 sản xuất là lớn nhất.

Ý nghĩa công thức Bayes

Các xác suất $P(H_1), ..., P(H_n)$ được xác định trước khi phép thử tiến hành và do đo chúng được gọi là *các xác suất tiên nghiệm*. Các xác suất $P(H_1|A), ..., P(H_n|A)$ được xác định sau khi phép thử được tiến hành và biến cố A đã xảy ra và do do chúng được gọi là *các xác suất hậu nghiệm*. Vì thế công thức Bayes còn được gọi là *công thức tính xác suất hậu nghiệm*.
* Công thức Bayes cho phép đánh giá lại xác suất xảy ra của các giả thiết sau khi đã biết kết quả của phép thử là biến cố A đã xảy ra.

Ví dụ 28

Tỉ lệ người dân ở tỉnh A nghiện thuốc lá là 30%, trong số người nghiện thuốc tỉ lệ người bị bệnh phổi là 60%. Trong số những người không nghiện thuốc có 20% bị bệnh phổi. Chọn ngẫu nhiên một người. Tính xác suất người đó nghiện thuốc trong hai trường hợp.
a) Biết người đó bị bệnh phổi.
b) Biết người đó không bị bệnh phổi.

Giải. Gọi H_1, H_2 là các biến cố người được chọn hút và không hút thuốc lá. Gọi A là biến cố người được chọn bị bệnh phổi.
Ta có: $\{H_1, H_2\}$ là nhóm đầy đủ. Theo công thức xác suất toàn phần:

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = 0,3 * 0,6 + 0,7 * 0,2 = 0,32$$

a. $P(H_1|A) = \dfrac{P(H_1)P(A|H_1)}{P(A)} = \dfrac{0,3 * 0,6}{0,32} = 9/16$
b. $P(H_1|\bar{A}) = \dfrac{P(H_1)P(\bar{A}|H_1)}{P(\bar{A})} = \dfrac{0,3 * (1 - 0,6)}{1 - 0,32} = 3/17$

Ví dụ 29

Một máy bay bắn độc lập 2 quả tên lửa vào một mục tiêu. Xác suất quả thứ nhất và thứ 2 trúng là 0,6 và 0,7. Nếu có 1 quả trúng thì mục tiêu bị tiêu diệt với xác suất 0,7 và nếu có 2 quả trúng thì xác suất này là 0,9.
a) Tính xác suất có ít nhất 1 quả trúng. b) Tính xác suất mục tiêu bị tiêu diệt.
c) Biết mục tiêu bị tiêu diệt, tính xác suất cả 2 quả đều trúng.

Giải. Gọi A_i là biến cố quả tên lửa thứ i trúng mục tiêu, $i = 1, 2$. Ta có A_1, A_2 độc lập.
a. $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1A_2) = 0,6 + 0,7 - 0,6 * 0,7 = 0,88$
b. Gọi H_i là biến cố có i quả tên lửa trúng, $i = 0, 1, 2$, và A là biến cố mục tiêu bị tiêu diệt. Ta có $\{H_0, H_1, H_2\}$ là nhóm đầy đủ.
$P(H_0) = P(\bar{A}_1\bar{A}_2) = 0,4 * 0,3 = 0,12; P(H_1) = P(A_1\bar{A}_2 \cup \bar{A}_1A_2) = 0,6 * 0,3 + 0,4 * 0,7 = 0,46; P(H_2) = P(A_1A_2) = 0,6 * 0,7 = 0,42.$
Theo công thức xác suất toàn phần:
$P(A) = P(H_0)P(A|H_0) + P(H_1)P(A|H_1) + P(H_2)P(A|H_2)$
$= 0,12 * 0 + 0,46 * 0,7 + 0,42 * 0,9 = 0,7$
c. $P(H_2|A) = \dfrac{P(H_2)P(A|H_2)}{P(A)} = \dfrac{0,42 * 0,9}{0,7} = 0,54$

Ví dụ 30

Người ta dùng một thiết bị để kiểm tra một loại sản phẩm nhằm xác định sản phẩm có đạt yêu cầu không. Biết rằng sản phẩm có tỉ lệ phế phẩm là p(%). Thiết bị có khả năng phát hiện đúng sản phẩm là phế phẩm với xác suất α và phát hiện đúng sản phẩm đạt chất lượng với xác suất β. Kiểm tra ngẫu nhiên một sản phẩm, tìm xác suất sao cho sản phẩm này:
a. Được kết luận là phế phẩm.
b. Được kết luận là đạt chất lượng thì lại là phế phẩm.
c. Được kết luận đúng với thực chất của nó.

Gợi ý. Gọi H_1, H_2 là biến cố sản phẩm được chọn là chính phẩm, phế phẩm. Gọi A là biến cố được kết luận là phế phẩm. Ta có $\{H_1, H_2\}$ là nhóm đầy đủ, $P(H_2) = p, P(A|H_2) = \alpha, P(\bar{A}|H_1) = \beta$.
a. $P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2) = (1-p)(1-\beta) + p\alpha$
b. $P(H_1|A)$
c. $P(AH_2) + P(\bar{A}H_1)$

8. Dãy phép thử Bernoulli

Định nghĩa

Dãy n phép thử được gọi là dãy n phép thử Bernoulli nếu thỏa mãn các điều kiện sau:
- Các phép thử độc lập
- Trong mỗi phép thử chỉ xảy ra một trong hai biến cố, kí hiệu A (thành công) và $\bar{A}$ (thất bại).
- Xác suất xuất hiện biến cố A là $p = P(A)$ không đổi trong các phép thử.

Công thức xác suất nhị thức

Kí hiệu $p_n(k)$ là xác suất có đúng k lần xuất hiện biến cố A trong dãy n phép thử Bernoulli. Khi đó:

$$p_n(k) = C_n^k p^k (1-p)^{n-k}, k = \overline{0, n}.$$

Công thức trên còn được gọi là công thức Bernoulli.

Ví dụ 31

Xác suất để 1 quả trứng đem ấp nở ra gà con là 0.85. Đem ấp 10 quả trứng. Tính xác suất để có đúng 8 quả nở ra gà con.

Giải. Ta có mô hình dãy phép thử Bernoulli với $n = 10; p = 0,85$.
Xác suất có đúng 8 quả nở ra gà con:

$$p_{10}(8) = C_{10}^{8} 0,85^8 0,15^2$$

Ví dụ 32

Tín hiệu thông tin được phát đi 3 lần độc lập nhau. Xác suất thu được mỗi lần là 0.4.
a) Tìm xác suất để nguồn thu nhận được thông tin đúng 2 lần.
b) Tìm xác suất để nguồn thu nhận được thông tin đó.
c) Nếu muốn xác suất thu được tin $\geq$ 0.99 thì phải phát đi ít nhất bao nhiêu lần.

Giải.

Ta có mô hình dãy phép thử Bernoulli với $n = 3; p = 0,4$.
a. Xác suất để nguồn thu nhận được thông tin đúng 2 lần:

$$p_3(2) = C_3^2 0,4^2 0,6^1$$

b. Xác suất để nguồn thu nhận được thông tin đó:

$$p_3(k \geq 1) = 1 - p_3(0) = 1 - C_3^0 0,4^0 0,6^3$$

c. Gọi n là số tín hiệu cần phát đi. Theo giả thiết:

$$p_n(k \geq 1) \geq 0,99 \Leftrightarrow 1 - p_n(0) \geq 0,99$$

$$\Leftrightarrow C_n^0 0,4^0 0,6^n \leq 0,01 \Leftrightarrow n \geq \frac{ln(0,01)}{ln(0,6)} = 9,015$$

Vậy, $n = 10$.

Ví dụ 33

Trên giá có 4 cây súng loại 1 và 6 cây súng loại 2. Một xạ thủ chọn ngẫu nhiên 1 cây súng và bắn 5 phát độc lập. Biết xác suất bắn trúng khi sử dụng súng loại 1 và 2 lần lượt là 0,6 và 0,3. Tính xác suất có đúng 2 viên đạn trúng đích.

Giải. Gọi H_i là biến cố xạ thủ chọn được súng loại $i, i = 1,2$ và A là biến cố xạ thủ bắn trúng 2 viên. Ta có $\{H_1, H_2\}$ là nhóm đầy đủ.
Theo công thức xác suất toàn phần:

$$P(A) = P(H_1)P(A|H_1) + P(H_2)P(A|H_2)$$

trong đó:

$$P(H_1) = 4/10 = 0,4; P(H_2) = 6/10 = 0,6$$

$$P(A|H_1) = p_5(2) = C_5^2 0,6^2 0,4^3; P(A|H_2) = p_5(2) = C_5^2 0,3^2 0,7^3$$

Ví dụ 34

Trong năm học vừa qua, ở trường đại học, tỉ lệ sinh viên thi trượt môn Toán là 34%, thi trượt môn Lý là 20%, và trong số các sinh viên trượt môn Toán, có 50% sinh viên trượt môn Lý. Phải chọn bao nhiêu sinh viên của trường này sao cho, với xác suất không bé hơn 99%, trong số đó có ít nhất một sinh viên đậu cả hai môn Toán và Lý.

Giải. Gọi T, L là các biến cố sinh thi trượt môn Toán, Lý tương ứng. Ta có:

$$P(T) = 0,34; P(L) = 0,2; P(L|T) = 0,5$$

Xác suất thi đậu cả 2 môn:

$$P(\bar{T}\bar{L}) = 1 - P(T \cup L) = 1 - [P(T) + P(L) - P(TL)]$$
$$= 1 - P(T) - P(L) + P(T)P(L|T) = 1 - 0,34 - 0,2 + 0,34 * 0,5 = 0,63$$

Gọi n là số sinh viên cần khảo sát. Ta có mô hình Bernoulli với n phép thử và xác suất $p = 0,63$. Theo giả thiết:

$$p_n(k \geq 1) = 1 - p_n(0) \geq 0,99 \Leftrightarrow (1 - 0,63)^n \leq 0,01 \Leftrightarrow n \geq 4,63 \Rightarrow n = 5$$

Số lần có khả năng lớn nhất

Bài toán: Tìm k sao cho xác suất $p_n(k)$ đạt giá trị lớn nhất.
Kết quả: Đặt $q = 1 - p$ và gọi K là giá trị cần tìm. Khi đó K là số nguyên thỏa điều kiện:

$$np - q \leq K \leq np - q + 1$$

Ví dụ 35

Một xưởng dệt có 50 máy dệt. Xác suất mỗi máy dệt bị hỏng trong mỗi ca làm việc là 0.1. Tìm số máy hỏng với khả năng lớn nhất trong 1 ca làm việc.

Giải. Gọi K là số máy hỏng với khả năng lớn nhất. Ta có mô hình Bernoulli với $n = 50; p = 0,1$.
Lại có: $q = 1 - p = 0,9$. Giá trị K thỏa điều kiện:

$$np - q \leq K \leq np - q + 1 \Leftrightarrow 4,1 \leq K \leq 5,1$$

Vậy, $K = 5$.

Phụ lục về mô phỏng

Ta kí hiệu R là giá trị ngẫu nhiên nằm trong đoạn [0,1] được phát sinh bởi hàm random trong máy tính.

Mô phỏng gieo đồng xu

- Phát sinh giá trị của R
- Nếu $R < 0.5$ thì cho kết quả xuất hiện mặt «sấp», ngược lại thì cho kết quả mặt «ngửa».

Mô phỏng gieo con xúc xắc

- Phát sinh giá trị của R
- Nếu R nằm trong khoảng $\left[\frac{i-1}{6}; \frac{i}{6}\right)$ thì cho xuất hiện kết quả là «mặt i», với $i = 1, 2, ..., 6$ tương ứng.