Thanh Son Trinh

Thống kê máy tính

08/2023

1 Xác suất có điều kiện

1.1 Lý thuyết

1.1.1 Định nghĩa

Xác suất có điều kiện

Giả sử điều kiện B có $P(B) > 0$, khi đó xác suất của sự kiện A biết rằng điều kiện B xảy ra, kí hiệu là $P(A \mid B)$ được định nghĩa là

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{1}$$

Hệ quả. Từ (1) ta được

$$P(A \cap B) = P(A \mid B) \cdot P(B). \tag{2}$$

Ta cũng có thể xem A là điều kiện và B là biến cố, khi đó

$$P(A \cap B) = P(B \mid A) \cdot P(A). \tag{3}$$

Tính chất

i. $0 \leqslant P(A \mid B) \leqslant 1$,

ii. $P(B \mid B) = 1$,

iii. $P(\overline{A} \mid B) = 1 - P(A \mid B)$,

iv. $P(A_1 \cup A_2 \mid B) = P(A_1 \mid B) + P(A_2 \mid B) - P(A_1 \cap A_2 \mid B)$.

Công thức xác suất có điều kiện mở rộng

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) \cdot \ldots P(A_n \mid A_1 \cap A_2 \ldots \cap A_{n-1}).$$

1.1.2 Sự độc lập và phụ thuộc của các biến cố

Hai biến cố A, B được gọi là độc lập khi việc xảy ra hay không của biến cố này không ảnh hưởng đến kết quả của biến cố kia. Nói cách khác, xác suất của biến cố A với điều kiện B bằng với xác suất của biến cố A khi không tính đến điều kiện B.

Hai biến cố độc lập

A, B được gọi là hai biến cố độc lập nếu

$$P(A \mid B) = P(A) \text{ hoặc } P(B \mid A) = P(B),$$

nghĩa là

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

Nếu hai biến cố không độc lập với nhau thì gọi là hai biến cố phụ thuộc. Do tính chất đối xứng nên nếu biến cố A phụ thuộc vào biến cố B thì biến cố B cũng phụ thuộc vào biến cố A.

- Nếu $P(A \mid B) > P(A)$ thì điều kiện B thuận lợi cho biến cố A.
- Nếu $P(A \mid B) < P(A)$ thì điều kiện B không thuận lợi cho biến cố A.

Nếu điều kiện B thuận lợi cho biến cố A thì điều kiện A cũng thuận lợi cho biến cố B.

1.1.3 Công thức Bayes

1. Công thức xác suất đầy đủ

Hệ đầy đủ

Hệ các biến cố $\{A_1, A_2, \ldots, A_n\}$ (với $n \in \mathbb{N}, n \geqslant 2$) của một phép thử được gọi là hệ đầy đủ nếu

i. $A_i \cap A_j = \varnothing$ với mọi $i \neq j$ (hay A_i, A_j xung khắc từng đôi một với nhau);

ii. $A_1 \cup A_2 \cup \ldots \cup A_n = \Omega$.

Đặc biệt, với mọi biến cố A thì hệ $\{A, \overline{A}\}$ đầy đủ.

Ví dụ 1. Một lớp có 4 tổ học sinh tham gia hội thi văn nghệ. Chọn ngẫu nhiên một tổ được giải nhất và gọi A_1, A_2, A_3, A_4 lần lượt là biến cố giải nhất thuộc về tổ 1, tổ 2, tổ 3 và tổ 4. Khi đó hệ $\{A_1, A_2, A_3, A_4\}$ là đầy đủ.

Nếu chưa biết xác suất của biến cố A nào đó, nhưng biết các xác suất $P(A_i)$ của một hệ đầy đủ $\{A_1, A_2, \ldots, A_n\}$ của không gian mẫu Ω và biết các xác suất có điều kiện $P(A \mid A_i)$ thì ta có thể dùng công thức sau, gọi là công thức xác suất đầy đủ, để tính xác suất của biến cố A:

Công thức xác suất đầy đủ

$$P(A) = \sum_{i=1}^{n} P(A \cap A_i) = \sum_{i=1}^{n} P(A_i) \cdot P(A \mid A_i). \tag{4}$$

Chú ý. Để áp dụng công thức xác suất đầy đủ, ta cần chỉ ra được một hệ biến cố đầy đủ. Ta thường gặp các trường hợp sau:

i. Công việc tiến hành trải qua 2 công đoạn. Thực hiện công đoạn thứ nhất có n khả năng xảy ra là các biến cố $A_1, A_2, \ldots, A_n$. Sau khi thực hiện công đoạn này, ta thực hiện công đoạn thứ 2 và biến cố A ở công đoạn này. Khi đó xác suất của biến cố A được tính theo công thức xác suất đầy đủ.

ii. Một tập hợp chứa n nhóm phần tử. Một nhóm phần tử có một tỉ lệ phần tử có tính chất **P** nào đó Lấy ngẫu nhiên từ tập hợp đó 1 phần tử. Gọi A_i là biến cố chọn được phần tử ở nhóm thứ i. Khi đó, xác suất chọn được phần tử có tính chất **P** được tính theo công thức xác suất đầy đủ.

Ví dụ 2. Một phân xưởng sản xuất lọ thuỷ tinh gồm 3 chiếc máy với tỉ lệ lần lượt là $20\%, 30\%$ và 50%. Xác suất lọ thuỷ tinh bị lỗi của các máy lần lượt là $0,001$; $0,005$; và $0,006$. Lấy ngẫu nhiên một lọ thuỷ tinh. Tìm xác suất để lấy ra sản phẩm lỗi.

Lời giải. Gọi A là biến cố "Lấy ra sản phẩm lỗi" và A_i là các biến cố "Lấy lọ thuỷ tinh sản xuất từ máy i" (với $i = \overline{1,3}$).

Dễ thấy hệ $\{A_1, A_2, A_3\}$ đầy đủ. Khi đó

$$\begin{aligned} P(A) &= P(A_1) \cdot P(A \mid A_1) + P(A_2) \cdot P(A \mid A_2) + P(A_3) \cdot P(A \mid A_3) \\ &= 0,2 \cdot 0,001 + 0,3 \cdot 0,005 + 0,5 \cdot 0,006 \\ &= 0,0065. \end{aligned}$$

2. Công thức Bayes

Công thức Bayes là công thức ngược, cho phép tính $P(B \mid A)$ khi biết xác suất có điều kiện $P(A \mid B)$ và một số thông tin khác. Dạng đơn giản nhất của công thức này là

$$P(B \mid A) = \frac{P(A \mid B) \cdot P(B)}{P(A)} \tag{5}$$

Công thức Bayes

Giả sử $\{A_1, A_2, \ldots, A_n\}$ là một hệ đầy đủ và A là một biến cố bất kỳ có thể xảy ra trong phép thử. Khi đó ta có công thức Bayes

$$P(A_k \mid A) = \frac{P(A_k) \cdot P(A \mid A_k)}{P(A)} = \frac{P(A_k) \cdot P(A \mid A_k)}{\sum\limits_{i=1}^{n} P(A_i) \cdot P(A \mid A_i)}, \tag{6}$$

với mọi $k = 1, 2, \ldots, n$.

Ví dụ 3. Một trang trại trồng hoa hướng dương phục vụ mùa Tết thu thập số liệu về việc hoa không đạt chất lượng. Trong đó, các nguyên nhân dẫn đến hoa không đạt chất lượng được chia làm ba nhóm: giống hoa, phân bón và các yếu tố tự nhiên (chẳng hạn như thời tiết, đất, nước,...) với xác suất tương ứng là $0,1; 0,6$ và $0,3$. Xác suất hoa không đạt chất lượng do giống hoa gây ra là $0,9$; do phân bón là $0,2$ và do các yếu tố tự nhiên là $0,5$. Nếu phân tích chất lượng hoa không đạt yêu cầu, nguyên nhân nào có khả năng cao nhất.

Lời giải. Gọi A là biến cố "Hoa không đạt chất lượng" và các biến cố

- A_1: "Hoa không đạt chất lượng do giống hoa"
- A_2: "Hoa không đạt chất lượng do phân bón"
- A_3: "Hoa không đạt chất lượng do các yếu tố tự nhiên"

Ta có hệ $\{A_1, A_2, A_3\}$ đầy đủ. Khi đó

(a) Xác suất để hoa không đạt chất lượng do giống hoa là

$$\begin{aligned} P(A_1 \mid A) &= \frac{P(A_1) \cdot P(A \mid A_1)}{P(A_1) \cdot P(A \mid A_1) + P(A_2) \cdot P(A \mid A_2) + P(A_3) \cdot P(A \mid A_3)} \\ &= \frac{0,1 \cdot 0,9}{0,1 \cdot 0,9 + 0,6 \cdot 0,2 + 0,3 \cdot 0,5} \\ &= \frac{1}{4}. \end{aligned}$$

(b) Xác suất để hoa không đạt chất lượng do phân bón là

$$\begin{aligned}\mathsf{P}(A_2 \mid A) &= \frac{\mathsf{P}(A_2)\cdot \mathsf{P}(A \mid A_2)}{\mathsf{P}(A_1)\cdot \mathsf{P}(A \mid A_1)+\mathsf{P}(A_2)\cdot \mathsf{P}(A \mid A_2)+\mathsf{P}(A_3)\cdot \mathsf{P}(A \mid A_3)}\\ &= \frac{0,6\cdot 0,2}{0,1\cdot 0,9+0,6\cdot 0,2+0,3\cdot 0,5}\\ &= \frac{1}{3}.\end{aligned}$$

(c) Xác suất để hoa không đạt chất lượng do các yếu tố khác là

$$\begin{aligned}\mathsf{P}(A_3 \mid A) &= \frac{\mathsf{P}(A_3)\cdot \mathsf{P}(A \mid A_3)}{\mathsf{P}(A_1)\cdot \mathsf{P}(A \mid A_1)+\mathsf{P}(A_2)\cdot \mathsf{P}(A \mid A_2)+\mathsf{P}(A_3)\cdot \mathsf{P}(A \mid A_3)}\\ &= \frac{0,3\cdot 0,5}{0,1\cdot 0,9+0,6\cdot 0,2+0,3\cdot 0,5}\\ &= \frac{5}{12}.\end{aligned}$$

Vậy hoa không đạt chất lượng nguyên nhân đến từ các yếu tố khác là cao nhất.

Ví dụ 4. Một shop hoa nhuộm màu hoa hồng để bán vào ngày 8/3 gồm hai loại xanh dương và xanh pastel với xác suất tương ứng là $0,85$ và $0,15$. Do điều kiện môi trường và các tác nhân khác nên $1/7$ số lượng hoa ở dung dịch xanh dương có màu thành xanh pastel, còn $1/8$ hoa ở dung dịch xanh pastel có màu xanh dương.

a. Tìm xác suất để shop thu được hoa màu xanh dương.

b. Giả sử hoa thu được có màu xanh dương. Hỏi xác suất thu được đúng hoa ở dung dịch nhuộm màu xanh dương.

Lời giải. Gọi các biến cố sau:

- A: "Hoa ở dung dịch xanh dương"
- B: "Hoa ở dung dịch xanh pastel"
- F_A: "Hoa màu xanh dương"
- F_B: "Hoa màu xanh pastel"

Ta có

$$\mathsf{P}(A) = 0,85; \mathsf{P}(B) = 0,15; \mathsf{P}(F_B \mid A) = 1/7; \mathsf{P}(F_A \mid B) = 1/8.$$

a. Áp dụng công thức xác suất đầy đủ, xác suất thu được hoa có màu xanh dương là

$$\begin{aligned}\mathsf{P}(F_A) &= \mathsf{P}(A)\cdot \mathsf{P}(F_A \mid A)+\mathsf{P}(B)\cdot \mathsf{P}(F_A \mid B)\\ &= 0,85\cdot \frac{6}{7}+0,15\cdot \frac{1}{8}\\ &\approx 0,7473.\end{aligned}$$

b. Áp dụng công thức Bayes ta có:

$$\begin{aligned}\mathsf{P}(A \mid F_A) &= \frac{\mathsf{P}(A)\cdot \mathsf{P}(F_A \mid A)}{\mathsf{P}(F_A)}\\ &= \frac{0,85\cdot \frac{6}{7}}{0,7473}\\ &\approx 0,975.\end{aligned}$$

Nhận xét. Công thức Bayes cho phép đánh giá lại xác suất xảy ra các biến cố A_k sau khi đã có thông tin về biến cố A. Lưu ý rằng, để có thể sử dụng công thức (4) hoặc (6) thì cần có hệ đầy đủ. Ngoài ra, nếu dùng (4) cho ta xác suất không có điều kiện thì (6) cho ta phép tính có điều kiện, trong đó biến cố A_k cần tính xác suất phải là một thành phần của hệ đầy đủ đang xét. Trong trường hợp không có hoặc khó xác định hệ đầy đủ, thì ta dùng công thức (5), khi đó việc tính $P(A)$ sẽ khó hơn là dùng công thức (4).

Ví dụ 5. Một chiếc máy gồm 2 động cơ hoạt động độc lập với nhau, với xác suất làm việc tốt trong một khoảng thời gian nào đó của mỗi động cơ là $0,95$ và $0,98$. Ở một thời điểm trong khoảng thời gian trên, người ta thấy máy ngừng làm việc (do có ít nhất một động cơ bị hỏng), tìm xác suất để chỉ có động cơ thứ hai bị hỏng.

Lời giải. Gọi A_1, A_2 lần lượt là các biến cố động cơ thứ nhất hoạt động tốt, động cơ thứ hai hoạt động tốt. Khi đó, xét các biến cố sau:

- B_1: "Cả hai động cơ hoạt động tốt" hay $B_1 = A_1 \cap A_2$;
- B_2: "Động cơ I hoạt động tốt và động cơ II bị hỏng" hay $B_2 = A_1 \cap \overline{A_2}$;
- B_3: "Động cơ I bị hỏng và động cơ II hoạt động tốt" hay $B_3 = \overline{A_1} \cap A_2$;
- B_4: "Cả hai động cơ đều bị hỏng" hay $B_4 = \overline{A_1} \cap \overline{A_2}$.

Ta có hệ $\{B_1, B_2, B_3, B_4\}$ là hệ đầy đủ. Do tính độc lập của các động cơ nên

$$\begin{aligned}
P(B_1) &= P(A_1) \cdot P(A_2) = 0,95 \cdot 0,98 = 0,931;\\
P(B_2) &= P(A_1) \cdot P(\overline{A_2}) = 0,95 \cdot 0,02 = 0,019;\\
P(B_3) &= P(\overline{A_1}) \cdot P(A_2) = 0,05 \cdot 0,98 = 0,049;\\
P(B_4) &= P(\overline{A_1}) \cdot P(\overline{A_2}) = 0,05 \cdot 0,02 = 0,001.
\end{aligned}$$

Gọi A là biến cố "Máy ngừng làm việc". Khi đó

$$P(A \mid B_1) = 0, P(A \mid B_2) = P(A \mid B_3) = P(A \mid B_4) = 1.$$

Áp dụng công thức (6) ta có xác suất cần tìm là

$$P(B_2 \mid A) = \frac{P(B_2) \cdot P(A \mid B_2)}{\sum\limits_{i=1}^{4} P(B_i) \cdot P(A \mid B_i)} = \frac{0,019}{0,019 + 0,049 + 0,001} = \frac{19}{69}.$$

Cách 2. (Sử dụng công thức (5)) Ta có

$$A = (A_1 \cap \overline{A_2}) \cup (\overline{A_1} \cap A_2) \cup (\overline{A_1} \cap \overline{A_2}).$$

Do tính xung khắc và độc lập của các biến cố trên nên ta có

$$P(A) = P(A_1) \cdot P(\overline{A_2}) + P(\overline{A_1}) \cdot P(A_2) + P(\overline{A_1}) \cdot P(\overline{A_2}) = 0,069.$$

Mặt khác, $B_2 \cap A = A_1 \cap \overline{A_2}$ nên $P(B_2 \cap A) = 0,019$. Từ đó suy ra kết quả mà không cần tìm hệ đầy đủ, tuy nhiên, việc tính $P(A)$ khó khăn hơn.

1.2 Vận dụng

Câu 1: Một hộp có 2 lá thăm được đánh số chẵn hoặc lẻ.

a. Tính xác suất để cả hai lá thăm đều là số lẻ.

b. Biết rằng lá thăm thứ nhất bốc ra là số lẻ. Tính xác suất để cả hai lá thăm đều là số lẻ.

Hướng dẫn. Giả sử điều kiện B có $P(B) > 0$, khi đó xác suất của sự kiện A biết rằng điều kiện B xảy ra, kí hiệu là $P(A \mid B)$ được định nghĩa là

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Câu 2: Một người tham gia thi 2 kỳ thi TOEIC Nghe - Đọc và Nói - Viết để lấy chứng chỉ xét tốt nghiệp. Biết rằng xác suất để đủ điểm Nghe - Đọc là $0,6$. Nếu đủ điểm Nghe - Đọc thì khả năng thí sinh đó đủ điểm Nói - Viết là $0,8$. Nếu Nghe - Đọc không đủ điểm thì khả năng thí sinh đó đủ điểm Nói - Viết chỉ là $0,6$. Tính xác suất trong các trường hợp sau:

a. Thí sinh đó chỉ đủ điểm một kỳ thi.

b. Thí sinh đó đủ điểm cả hai kỳ thi.

Hướng dẫn. $P(A \cap B) = P(A) \cdot P(B \mid A)$

Câu 3: Một lô hàng có 100 sản phẩm, trong đó có 90 sản phẩm tốt và 10 sản phẩm lỗi. Kiểm tra ngẫu nhiên liên tiếp không hoàn lại 5 sản phẩm. Nếu có ít nhất 1 sản phẩm lỗi trong 5 sản phẩm được kiểm tra thì không nhận lô hàng. Tìm xác suất để nhận lô hàng.

Hướng dẫn. Công thức xác suất có điều kiện mở rộng

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \cdot P(A_2 \mid A_1) \cdot P(A_3 \mid A_1 \cap A_2) \cdot \ldots P(A_n \mid A_1 \cap A_2 \ldots \cap A_{n-1}).$$

——— HẾT ———