Bài 4:
Xác suất có điều kiện

Nội dung

- Định nghĩa xác suất có điều kiện
- Tính
  - xác suất từ phân hoạch
  - xác suất có điều kiện bằng công thức Bayes
- Ứng dụng: xích Markov

Định nghĩa

- Cần xem xét sự thay đổi xác suất của một biến cố A khi biến cố B đã xảy ra trước đó.
- Xác suất của biến cố A trong trường hợp này được gọi ***là xác suất có điều kiện (conditional probability) của biến cố A khi biết biến cố B xảy ra*** – ký hiệu là **Pr(A|B)**
- **Đ/n**: nếu A và B là 2 biến cố với Pr(B) > 0 thì

$$\Pr(A \mid B) = \frac{\Pr(AB)}{\Pr(B)}$$

Ví dụ

1 hộp gồm 8 bi trắng, 2 bi đỏ. Lần lượt bốc từng bi. Giả sử lần đầu tiên bốc được bi trắng. Xác định xác suất lần thứ 2 bốc được bi đỏ.

B (biến cố đã xảy ra trước) là biến cố lần 1 bốc được bi trắng. Gọi A là biến cố lần 2 bốc được bi đỏ.

Xác suất lần 2 bốc được bi đỏ là Pr(A|B) (xác suất của A với điều kiện B)

$$\Pr(A \mid B) = \frac{\Pr(AB)}{\Pr(B)} = \frac{\frac{8}{10} \times \frac{2}{9}}{\frac{8}{10}} = \frac{2}{9}$$

với Pr(AB)=xác suất lần 1 bốc bi trắng và lần 2 bốc bi đỏ, Pr(B)=xác suất lần 1 bốc bi trắng.

Tính chất

- Pr(A|B) = Pr(A) ⇔ A và B là hai biến cố độc lập (independent events).
- Luật nhân (multiplication rule):
  - Pr(AB) = Pr(A).Pr(B|A)
  - Pr(AB) = Pr(B).Pr(A|B)
- $Pr(A_1 \ldots A_n) = Pr(A_1) \times Pr(A_2|A1) \times Pr(A_3|A_1A_2) \times \ldots \times Pr(A_n|A_1A_2\ldots A_{n-1})$

  với $Pr(A_1 \ldots A_n) > 0$

Ví dụ

1 hộp gồm 8 bi trắng, 2 bi đỏ. Chọn 1 viên bi, sau đó bỏ lại viên bi đó vào hộp. Lặp lại bằng cách chọn 1 viên bi, sau đó lại bỏ viên bi đó vào hộp. Giả sử lần đầu tiên bốc được bi trắng. Xác định xác suất lần thứ 2 bốc được bi đỏ.

- B (biến cố đã xảy ra trước) là biến cố lần 1 bốc được bi trắng. Gọi A là biến cố lần 2 bốc được bi đỏ.

$$\Pr(A \mid B) = \frac{2}{10} = \Pr(A)$$

A, B là 2 biến cố độc lập

Ví dụ

- 1 hộp gồm 8 bi trắng, 2 bi đỏ. Lần lượt bốc từng bi. Xác định xác suất lần thứ 1, thứ 2 bốc được bi trắng và lần thứ 3 bốc được bi đỏ.
- Giải: Gọi A_i là biến cố lần thứ i bốc được bi trắng. (=> A_i^c là biến cố lần thứ i bốc được bi đỏ)

  Xác suất lần thứ 1, 2 bốc được bi trắng, lần thứ 3 bốc được bi đỏ là:

$$\Pr(A_1 A_2 A_3^c) = \Pr\left(A_1\right)\Pr\left(A_2 \mid A_1\right)\Pr\left(A_3^c \mid A_1 A_2\right) = \frac{8}{10}\times\frac{7}{9}\times\frac{2}{8} = \frac{7}{45}$$

Tính xác suất bằng phân hoạch (total probability formula)

- Không gian mẫu S được hợp thành từ các biến cố A_i (i=1…k) tách rời – một **phân hoạch** của S

$A_i \cap A_j = \emptyset$ (i≠j)

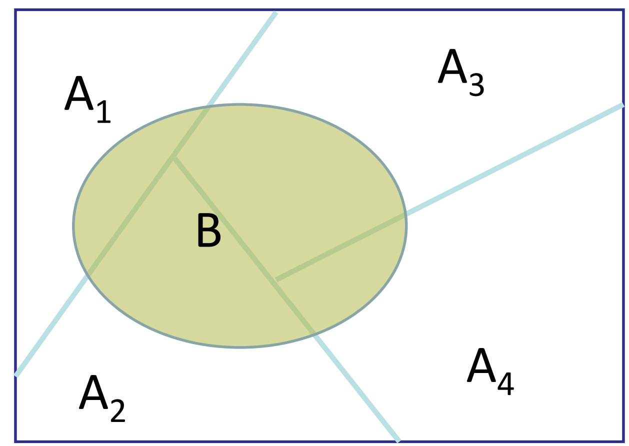
$A_1 \cup A_2 \cup \ldots \cup A_k = S$

- Không gian mẫu S được hợp thành từ các biến cố A_i (i=1…k) tách rời – một **phân hoạch** của S

$A_i \cap A_j = \emptyset$ (i≠j)

$A_1 \cup A_2 \cup \ldots \cup A_k = S$ với $\Pr(A_j) > 0$ (j=1…k)

- Xem hình vẽ



$$\Pr(B) = \sum_{j=1}^{k} \Pr(A_j B) = \sum_{j=1}^{k} \Pr(A_j)\Pr(B \mid A_j)$$

Ví dụ

- 1 hộp gồm 8 bi trắng, 2 bi đỏ. Lần lượt bốc từng bi. Xác định xác suất lần thứ 2 bốc được bi đỏ.
- Giải: Gọi A_i là biến cố lần thứ i bốc được bi trắng. (=> A_i^c là biến cố lần thứ i bốc được bi đỏ)

$\{A_1, A_1^c\}$ là 1 phân hoạch của S vì $A_1 \cap A_1^c = \emptyset$ và $A_1 \cup A_1^c = S$

$$\Pr(A_2^c) = \Pr(A_1 A_2) + \Pr(A_1^c A_2)$$
$$= \Pr(A_1)\Pr(A_2^c \mid A_1) + \Pr(A_1^c)\Pr(A_2^c \mid A_1^c)$$
$$= \frac{8}{10} \times \frac{2}{9} + \frac{2}{10} \times \frac{1}{9} = \frac{18}{90}$$

Ví dụ

- Đi xét nghiệm máu, kết quả dương tính. Có bị bệnh không?
- Kinh nghiệm cho biết
  - trong 10000 người chỉ có 1 người bị bệnh
  - nếu một người bị bệnh thì xác suất xét nghiệm ra dương tính là 90%
  - nếu một người *không* bị bệnh thì xác suất ra dương tính là 10%
- Nhận xét: không thể dùng phương pháp đếm

Định lý Bayes

- Đặt
  - A = “bị bệnh”
  - B = “dương tính”

Cần tính Pr(A|B)

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \times \Pr(A)}{\Pr(B \mid A) \times \Pr(A) + \Pr(B \mid A^c) \times \Pr(A^c)}$$

$$= \frac{0.9 \times 0.0001}{0.9 \times 0.0001 + 0.1 \times 0.9999} = 0.00090.$$

- Tức là xác suất bị bệnh khi xét nghiệm dương tính là 0.0009

Định lý Bayes

- **Công thức Bayes**: *Giả sử các biến cố $A_1, \ldots, A_k$ hình thành một phân hoạch của không gian S và $Pr(A_j) > 0$ ($\forall j = 1, \ldots, k$), B là một biến cố bất kỳ thỏa $Pr(B) > 0$ thì, $\forall i = 1, \ldots, k$,*

$$\Pr(A_i \mid B) = \frac{\Pr(A_i) \times \Pr(B \mid A_i)}{\Pr(B)} = \frac{\Pr(A_i) \times \Pr(B \mid A_i)}{\sum_{j=1}^{k} \Pr(A_j) \times \Pr(B \mid A_j)}.$$

  - *Xác suất tiên nghiệm (priori probability) : $Pr(A_i)$*
  - *Xác suất hậu nghiệm (posterior probability): $Pr(A_i|B)$*

XS tiên nghiệm – XS hậu nghiệm

- Xác suất tiên nghiệm (priori probability): xác suất được biết trước khi thực hiện thí nghiệm.
- Xác suất hậu nghiệm (posterior probability): xác suất được xác định sau khi thực hiện thí nghiệm.

Tính xác suất hậu nghiệm

- Dựa vào định lý Bayes
- Nếu 1 thí nghiệm tiến hành qua nhiều giai đoạn: ***xác suất hậu nghiệm ở giai đoạn trước là xác suất tiên nghiệm của giai đoạn tiếp theo***
- Ví dụ (bài toán xét nghiệm): Nếu xét nghiệm lần 2, kết quả dương tính.

$$\Pr(A)=0.00090 \Rightarrow \Pr\left(A \mid B\right)=\frac{\Pr(B \mid A)\times \Pr(A)}{\Pr(B \mid A)\times \Pr(A)+\Pr(B \mid A^c)\times \Pr(A^c)}$$

$$=\frac{0.9\times 0.0009}{0.9\times 0.0009+0.1\times\left(1-0.0009\right)}=0.008$$

Tiến trình ngẫu nhiên

- **<u>Ví dụ</u>**: có 5 đường dây điện thoại, cứ 2 phút đếm số lượng đường dây bị bận
  - X_i: số đường dây bị bận ở thời điểm thứ i = 1…n…,
- Chuỗi $X_1, X_2, \ldots, X_n, \ldots$ được gọi là một ***tiến trình ngẫu nhiên với tham số thời gian rời rạc.***
- Mô hình xác suất được thể hiện bởi

$$\Pr(X_{n+1} = x_{n+1} \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

với mọi n > 1

  - Mẫu là liên tục

Xích Markov

- **Xích Markov**: là một tiến trình ngẫu nhiên với

$$\Pr(X_{n+1} = x_{n+1} \mid X_1 = x_1, \ldots, X_n = x_n)$$
$$= \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n).$$

- **Xích Markov hữu hạn**: tại mỗi thời điểm, xích chỉ được nhận 1 trong k trạng thái $s_1, \ldots, s_k$.
- **Xác suất chuyển** (1 bước) từ trạng thái s_i ở thời điểm n đến s_j ở thời điểm n+1 là

$$\Pr(X_{n+1} = s_j | X_n = s_i)$$

Xích Markov

- **Xích Markov có xác suất chuyển không đổi:**

$$\Pr(X_{n+1} = s_j \mid X_n = s_i) = p_{ij}, \qquad \forall n = 1,2,\ldots$$

- **Chuyển 2 bước**

$$\Pr(X_{n+2} = s_j \mid X_n = s_i) = p_{ij}^{(2)} = \sum_{r=1}^{k} p_{ir} \times p_{rj}$$

- **Ma trận 1-bước chuyển**

$$P = \begin{bmatrix} p_{11} & \cdots & p_{1k} \\ \cdots & \cdots & \cdots \\ p_{k1} & \cdots & p_{kk} \end{bmatrix}.$$

→ Ma trận m-bước chuyển là P^m

Ví dụ

- Trở lại ví dụ đường dây điện thoại, ma trận 1-bước chuyển

$$P = \begin{array}{c|cccccc} & b_0 & b_1 & b_2 & b_3 & b_4 & b_5 \\ \hline b_0 & 0.1 & 0.4 & 0.2 & 0.1 & 0.1 & 0.1 \\ b_1 & 0.2 & 0.3 & 0.2 & 0.1 & 0.1 & 0.1 \\ b_2 & 0.1 & 0.2 & 0.3 & 0.2 & 0.1 & 0.1 \\ b_3 & 0.1 & 0.1 & 0.2 & 0.3 & 0.2 & 0.1 \\ b_4 & 0.1 & 0.1 & 0.1 & 0.2 & 0.3 & 0.2 \\ b_5 & 0.1 & 0.1 & 0.1 & 0.1 & 0.4 & 0.2 \end{array}$$

Xác suất chuyển trạng thái từ 1 đường dây bận tại thời điểm n sang 4 đường dây bận tại thời điểm kế tiếp (n+1) là: $P_{b1,b4}=0.1$

NX: tổng dòng = 1

Ví dụ

- Ma trận 2 bước chuyển

$$P^2 = P \times P = \begin{array}{c|cccccc} & b_0 & b_1 & b_2 & b_3 & b_4 & b_5 \\ \hline b_0 & 0.14 & 0.23 & 0.20 & 0.15 & 0.16 & 0.12 \\ b_1 & 0.13 & 0.24 & 0.20 & 0.15 & 0.16 & 0.12 \\ b_2 & 0.12 & 0.20 & 0.21 & 0.18 & 0.17 & 0.12 \\ b_3 & 0.11 & 0.17 & 0.19 & 0.20 & 0.20 & 0.13 \\ b_4 & 0.11 & 0.16 & 0.16 & 0.18 & 0.24 & 0.15 \\ b_5 & 0.11 & 0.16 & 0.15 & 0.17 & 0.25 & 0.16 \end{array}$$

- Xác suất chuyển trạng thái từ 1 đường dây bận tại thời điểm n sang 4 đường dây bận tại thời điểm (n+2) là: $P_{b1,b4}$=0.16

Xích Markov

- Tại thời điểm đầu, đặt $v_i = \Pr(X_1 = s_i)$ với i = 1…k thì

  $V_1 = (v_1,\ldots,v_k)$ là **vector xác suất đầu**
- Tại thời điểm n > 1

$$\Pr(X_n = s_j) = \sum_{i=1}^{k} \Pr(X_1 = s_i \wedge X_n = s_j)$$

$$= \sum_{i=1}^{k} \Pr(X_1 = s_i)\Pr(X_n = s_j \mid X_1 = s_i)$$

$$\Rightarrow V_n = \begin{bmatrix} \Pr(X_n = s_1) \\ \vdots \\ \Pr(X_n = s_k) \end{bmatrix}^T = V_1 \mathbf{P}^{n-1}$$

Ví dụ

- Trở lại ví dụ đường dây điện thoại, ma trận 1-bước chuyển

$$P = \begin{array}{c|cccccc} & b_0 & b_1 & b_2 & b_3 & b_4 & b_5 \\ \hline b_0 & 0.1 & 0.4 & 0.2 & 0.1 & 0.1 & 0.1 \\ b_1 & 0.2 & 0.3 & 0.2 & 0.1 & 0.1 & 0.1 \\ b_2 & 0.1 & 0.2 & 0.3 & 0.2 & 0.1 & 0.1 \\ b_3 & 0.1 & 0.1 & 0.2 & 0.3 & 0.2 & 0.1 \\ b_4 & 0.1 & 0.1 & 0.1 & 0.2 & 0.3 & 0.2 \\ b_5 & 0.1 & 0.1 & 0.1 & 0.1 & 0.4 & 0.2 \end{array}$$

- Vector xác suất đầu $V_1 = (0.5,0.3,0.2,0,0,0)$

Ví dụ trên

- Hỏi: sau 4 phút, xác suất để có 3 đường dây bận là bao nhiêu?
- Tính

  4 phút = 2 × 2 phút => chuyển trạng thái qua 2 bước

  $V_3 = V_1.P^2 = (0.13,0.23,0.20,0.16,0.16,0.12)$.

→ $Pr(X_3 = 3) = 0.16$

Tóm tắt

- Nội dung chính
  - Xác suất có điều kiện
  - Công thức xác suất tổng
  - Công thức Bayes
  - Xích Markov
- Từ khóa
  - Xác suất có điều kiện (conditional probaility),
  - định lý Bayes (Bayes's theorem),
  - xích Markov (Markov chain)

Ví dụ

- Hai tên cướp bịt mặt tấn công 1 ngân hàng. Tuy nhiên, người thu ngân đã kịp thời nhấn chuông báo động và khóa cửa ra vào. Các tên cướp nhận ra rằng chúng đã mắc kẹt, vì vậy quyết định cởi mặt nạ và trà trộn vào đám đông. Đối mặt với 40 người trong ngân hàng đều tự nhận mình vô tội, cảnh sát quyết định sử dụng máy phát hiện nói dối. Giả sử xác suất 1 người phạm tội bị máy phát hiện nói dối báo động là 0.85, xác suất 1 người vô tội bị máy báo động 0.08. Tính xác suất ông Smith là 1 trong 2 tên cướp biết rằng ông Smith bị máy phát hiện nói dối báo động.

Ví dụ (tt)

A: là tên cướp

B: bị máy phát hiện nói dối báo động

$Pr(A) = 2/40$

$Pr(B|A) = 0.85$, $Pr(B|A^c) = 0.08$

Tính $Pr(A|B)$

Ví dụ

- Bài tập thực hành 1:
- Dùng R để sinh ra bảng kết quả ở trang 26 (giáo trình KHNT)
- Dùng R để mô phỏng bài tập xét nghiệm máu