THỐNG KÊ MÁY TÍNH

(Computational Statistics)

Trường Đại học Nha Trang

Khoa Công nghệ thông tin

Bộ môn Hệ thống thông tin

Giảng viên: TS.Nguyễn Khắc Cường

CHƯƠNG 4

XÁC SUẤT

4.3.Xác suất

- Định nghĩa
  - Theo quan điểm cổ điển
    - Xét biến cố A có n(A) biến cố sơ cấp
    - Và không gian biến cố có n(Ω) biến cố cùng khả năng xuất hiện
    - Thì $P(A) = \frac{n(A)}{n(\Omega)}$ là xác suất của biến cố A

    → Nhận xét:
      - Ước lượng xác suất mà không cần thực hiện phép thử
      - Khó xác định được n(Ω)
  - Theo quan điểm thống kê
    - Thực hiện lặp đi lặp lại n phép thử ngẫu nhiên
    - Quan sát: biến cố A xảy ra f lần

4.3.Xác suất

- Định nghĩa
  - Theo quan điểm thống kê
    - Nếu tăng số lần thực hiện n lên đủ lớn → thì xác suất xảy ra biến cố A được ước lượng là f/n

$$P(A) = \lim_{n\to\infty} \frac{f}{n} \cong \frac{f}{n}$$

- Dùng R tính xác suất (cổ điển)

Xác định biến cố sơ cấp của không gian mẫu

```
> tosscoin(3)
  toss1 toss2 toss3
1     H     H     H
2     T     H     H
3     H     T     H
4     T     T     H
5     H     H     T
6     T     H     T
7     H     T     T
8     T     T     T
```

Xác định xác suất biến cố sơ cấp của không gian mẫu

```
> S<-tosscoin(3, makespace=TRUE)
> S
  toss1 toss2 toss3 probs
1     H     H     H 0.125
2     T     H     H 0.125
3     H     T     H 0.125
4     T     T     H 0.125
5     H     H     T 0.125
6     T     H     T 0.125
7     H     T     T 0.125
8     T     T     T 0.125
```

4.3.Xác suất

- Dùng R tính xác suất
  - Hàm xác định xác suất các biến cố: Prob(<biến cố>)
  - VD:

```
> library(prob)
> S<-tosscoin(3, makespace=TRUE)
> S
  toss1 toss2 toss3 probs
1     H     H     H 0.125
2     T     H     H 0.125
3     H     T     H 0.125
4     T     T     H 0.125
5     H     H     T 0.125
6     T     H     T 0.125
7     H     T     T 0.125
8     T     T     T 0.125
```

```
> S<-tosscoin(3, makespace=TRUE)
> A<-subset(S, toss1=="H" & toss2=="H")
> prob(A)
'prob' is deprecated; use 'Prob' instead.
[1] 0.25
>
>
> Prob(A)
[1] 0.25
```

```
> A<-subset(S, toss1=="T" & toss2=="H" & toss3=="H")
> Prob(A)
[1] 0.125
```

```
> A<-subset(S, toss1=="H")
> Prob(A)
[1] 0.5
```

4.3.Xác suất

- Các tính chất cơ bản của xác suất
  - Hệ tiên đề Kolmogorov
    - Xét
      - S : Không gian mẫu
      - P(A) : xác suất của biến cố A
    - Tiên đề:
      - (1) $P(A) \geq 0, \forall A \subseteq S$
      - (2) $P(S) = 1$
      - (3) Nếu $A_1, A_2, \ldots, A_n$ xung khắc thì $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i), \forall n$
  - Tính chất khác
    - Xét A, B là 2 biến cố bất kỳ thuộc không gian biến cố S

$P(\emptyset) = 0$ $\quad 0 \leq P(A) \leq 1$ $\quad P(A^c) = 1 - P(A)$

Nếu $A \subset B$ thì $P(A) \leq P(B)$ $\quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P\left(A_i \cap A_j\right) + .. + (-1)^{n-1} P(\bigcap_{i=1}^{n} A_i)$$

$\{A1, A2, .., An\}$ hệ đầy đủ, thì $P(B) = \sum_{i=1}^{n} P(B \cap A_i)$

4.4.Xác suất có điều kiện

- Định nghĩa
  - Xác suất xảy ra biến cố A khi biến cố B đã xảy ra

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

- Qui tắc nhân
  - Nếu các biến cố Ai, $\mathrm{i} = 1..\mathrm{n}$ là độc lập thì

$$P(\cap_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i)$$

  - Nếu các biến cố Ai, $\mathrm{i} = 1..\mathrm{n}$ không độc lập thì

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)P(A_n|A_1A_{2\ldots}A_{n-1})$$

4.4.Xác suất có điều kiện

- Qui tắc xác suất đầy đủ
  - Xét không gian mẫu M
  - $H_1, H_2, \ldots, H_n$ là một phân hoạch của M
    - $H_i \cap H_j = \varnothing$ với mọi $i \neq j$ và $\bigcup_{i=1}^{n} A_i = \Omega$
  - A là biến cố bất kỳ có liên quan đến phân hoạch này
  - Xác suất của A có thể xác định bằng công thức xác suất đầy đủ

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

Trong đó

- $P(H_i)$:các xác suất tiên nghiệm (prior probability) của A (xác suất của giả thiết)
- $P(A \mid H_i)$: xác suất khả dĩ (likelihood probability)

- VD:
  - Chọn sản phẩm chính phẩm >>>

4.4.Xác suất có điều kiện

- Công thức Bayes
  - Xét không gian mẫu M
  - $H_1, H_2, \ldots, H_n$ là một phân hoạch của M
    - $H_i \cap H_j = \varnothing$ với mọi $i \neq j$ và $\bigcup_{i=1}^{n} A_i = \Omega$
  - A là biến cố bất kỳ có liên quan đến phân hoạch này
  - Xác suất hậu nghiệm của H_i được xác định bằng công thức

$$P(H_i|A) = \frac{P(H_k)\,P(A|H_k)}{\sum_{i=1}^{n} P(H_i)P(A|H_i)} \qquad k = 1,2,\ldots,n$$

Trong đó

- $P(H_i \mid A)$: các xác suất hậu nghiệm (Posterior probability) của các H_i
- $P(H_i)$:các xác suất tiên nghiệm (prior probability) của A (xác suất của giả thiết)
- $P(A \mid H_i)$: xác suất khả dĩ (likelihood probability)
- $P(A) = \sum_{i=1}^{n} P(H_i)P(A|H_i)$

VD: Chọn sản phẩm từ nhà máy >>>

Q / A