

Chương 3: RA QUYẾT ĐỊNH (DECISION MAKING)

1. Các tổn thất và ra quyết định (losses and decision making)

Hàm tổn thất

Định nghĩa 1. Hàm tổn thất là một hàm số từ L từ $D \times \Theta$ vào $[0, +\infty]$.

Hàm tổn thất đánh giá mức phạt (sai số) $L(d, \theta)$ khi quyết định d với tham số nhận giá trị θ .

Một nền tảng cơ bản của lý thuyết quyết định Bayes là suy luận thống kê được bắt đầu từ việc xác định ba nhân tố:

- Họ các phân phối xác suất của các quan sát $f(x|\theta)$
- Phân phối tiên nghiệm của các tham số $\pi(\theta)$
- Quyết định với hàm tổn thất tương ứng $L(d, \theta)$.

Ra quyết định

Từ quan điểm của lý thuyết quyết định, mô hình thống kê được xác định bởi ba không gian: X là không gian các quan sát, Θ là không gian các tham số và D là không gian các quyết định. Các quyết định có vai trò quan trọng trong bài toán phân tích rủi ro.

Một số hàm tổn thất (loss function) thông dụng

Hàm tổn thất là sai số giữa giá trị đúng (không biết) θ và giá trị ước lượng $\hat{\theta}$, thông thường có 3 dạng hàm phổ biến là:

- a. Hàm tổn thất dạng bình phương (quadratic loss)

$$L(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$$

- b. Hàm tổn thất dạng trị tuyệt đối

$$L(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$$

- c. Hàm tổn thất dạng 0-1

$$L(\hat{\theta}, \theta) = \begin{cases} 0 & \text{nếu } \hat{\theta} = \theta \\ c & \text{nếu } \hat{\theta} \neq \theta \end{cases}$$

Trong ước lượng điểm của thống kê Bayes:

- Nếu chọn hàm tổn thất dạng bình phương thì ước lượng điểm Bayes tối ưu là trung bình của phân phối hậu nghiệm
- Nếu chọn hàm tổn thất dạng trị tuyệt đối thì ước lượng điểm Bayes tối ưu chính là trung vị của phân phối hậu nghiệm
- Nếu chọn hàm tổn thất dạng 0-1 thì ước lượng điểm Bayes tối ưu chính là số yếu vị của

phân phối hậu nghiệm.

Độ chính xác trong ước lượng Bayes

Giả sử phân phối hậu nghiệm $\pi(\theta|x)$, ước lượng điểm $\delta^\pi(x)$ của $h(\theta)$. Khi đó, độ chính xác của ước lượng được tính thông qua sai số bình phương hậu nghiệm (the posterior squared error):

$$E^\pi \left[(\delta^\pi(x) - h(\theta))^2 | x \right]$$

Nhận giá trị bằng $\text{var}^\pi(h(\theta)|x)$ khi $\delta^\pi(x) = E^\pi[h(\theta)|x]$.

2. Bài toán kiểm định giả thuyết trong thống kê Bayes

Giả sử mẫu ngẫu nhiên x_1, x_2, \dots, x_n bao gồm các quan sát độc lập và cùng tuân theo một phân phối xác suất $f(x)$, trong đó tham số của phân phối xác suất tuân theo là θ , nhận các giá trị trong không gian các tham số Θ .

Xét một tập con các giá trị của Θ là Θ_0 , tương ứng với giả thuyết (the null hypothesis) trong bài toán kiểm định giả thuyết (the hypothesis testing):

$$H_0: \theta \in \Theta_0$$

Suy ra, các giá trị $\Theta_1 = \Theta \setminus \Theta_0$ tương ứng với đối thuyết H_1 trong bài toán kiểm định giả thuyết. Khi đó, bài toán kiểm định giả thuyết là trường hợp riêng của bài toán lựa chọn mô hình (the model choice problem). Các quyết định trong bài toán kiểm định giả thuyết được dựa vào nhân tố Bayes.

Nhân tố Bayes

Định nghĩa 2. Nhân tố Bayes (the Bayes factor) là tỷ lệ giữa xác suất hậu nghiệm của giả thiết và đối thiết so với tỷ lệ của xác suất tiên nghiệm của giả thiết và đối thiết, tức là

$$B_{01}^\pi = \frac{P(x_1, x_2, \dots, x_n | \theta \in \Theta_0)}{P(x_1, x_2, \dots, x_n | \theta \in \Theta_1)} = \frac{\frac{P(\theta \in \Theta_0 | x_1, x_2, \dots, x_n)}{P(\theta \in \Theta_1 | x_1, x_2, \dots, x_n)}}{\frac{\pi(\theta \in \Theta_0)}{\pi(\theta \in \Theta_1)}}$$

- Nếu $B_{01}^\pi < 1$ thì bác bỏ H_0
- Nếu $B_{01}^\pi > 1$ thì chưa có cơ sở bác bỏ H_0

Trong trường hợp đơn giản nhất, $\Theta_0 = \{\theta_0\}$ và $\Theta_1 = \{\theta_1\}$, khi đó nhân tố Bayes chính là tỷ số hai hàm hợp lý với các giá trị của tham số chỉ nhận lần lượt một giá trị:

$$B_{01}^\pi = \frac{f(x_1, x_2, \dots, x_n | \theta_0)}{f(x_1, x_2, \dots, x_n | \theta_1)}$$

Trong hầu hết các trường hợp, nhân tố Bayes phụ thuộc vào thông tin tiên nghiệm.

Bài toán kiểm định hai phía

Bài toán kiểm định

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \\ \text{Mức ý nghĩa } \alpha \end{cases}$$

Tính miền mật độ xác suất cao nhất với xác suất $(1 - \alpha)$.

- Nếu miền mật độ xác suất cao nhất chứa giá trị θ_0 thì chưa có cơ sở bác bỏ H_0 , chấp nhận $\theta = \theta_0$.
- Nếu miền mật độ xác suất cao nhất không chứa giá trị θ_0 thì bác bỏ H_0 , tức là $\theta \neq \theta_0$.

Bài toán kiểm định một phía

Bài toán kiểm định

$$\begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases} \text{ hoặc } \begin{cases} H_0: \theta \geq \theta_0 \\ H_1: \theta < \theta_0 \end{cases}$$

Tính giá trị nhân tố Bayes và đưa ra quyết định bác bỏ H_0 hay không.

3. Suy luận trung bình của một tổng thể tuân theo phân phối chuẩn

Giả sử mẫu ngẫu nhiên bao gồm các quan sát x_1, x_2, \dots, x_n độc lập và cùng tuân theo một phân phối xác suất, với hàm mật độ xác suất là $f(x_i|\theta)$.

$$f(x_i|\theta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}}$$

Khi đó, hàm hợp lý có dạng (Lindley, 2011):

$$l(x_1, x_2, \dots, x_n|\theta) = \prod_{i=1}^n f(x_i|\theta) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\theta)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\sum_{i=1}^n (x_i-\theta)^2}{2\sigma^2}}.$$

Các phân tích dựa vào phân phối hậu nghiệm được tính toán dựa vào các trường hợp thông tin tiên nghiệm:

Trường hợp tiên nghiệm đều

Giả sử phân phối tiên nghiệm cho tham số trung bình của tổng thể θ có dạng

$$\pi(\theta) \propto 1$$

Suy ra phân phối hậu nghiệm cho tham số trung bình tổng thể θ có dạng

$$\pi(\theta|x_1, x_2, \dots, x_n) \propto e^{-\frac{\sum_{i=1}^n (x_i-\theta)^2}{2\sigma^2}} = e^{-\frac{\sum_{i=1}^n (x_i-\bar{X})^2 + n(\bar{X}-\theta)^2}{2\sigma^2}} \propto e^{-\frac{(\theta-\bar{X})^2}{2\left(\frac{\sigma}{\sqrt{n}}\right)^2}}.$$

Khi đó, phân phối hậu nghiệm của tham số trung bình tổng thể θ tuân theo phân phối chuẩn

với trung bình $\bar{X} = \frac{x_1+x_2+\dots+x_n}{n}$ và phương sai $\frac{\sigma^2}{n}$.

Ví dụ 1. Giả sử mẫu ngẫu nhiên gồm 10 quan sát từ phân phối chuẩn $N(\mu, \sigma^2)$ với giả định đã biết phương sai $\sigma^2 = 4$ với mẫu ngẫu nhiên

3.07 7.51 5.95 6.83 8.80 4.19 7.44 7.06 9.67 6.89

Giả sử tiên nghiệm cho tham số trung bình tổng thể $\theta = \mu$ là phân phối đều. Có thông tin cho rằng trung bình của tổng thể bằng 7. Hãy kiểm định lại thông tin trên với mức ý nghĩa 5%.

Giải:

Phân phối hậu nghiệm của tham số trung bình tổng thể $\theta = \mu$ là phân phối chuẩn với trung bình

$$\bar{X} = 6.741, \sigma'^2 = \frac{4}{10} = 0.4$$

Kiểm định hai phía về trung bình tổng thể

$$\begin{cases} H_0: \theta = 7 \\ H_1: \theta \neq 7 \end{cases}$$

Khoảng ước lượng tham số trung bình tổng thể $\theta = \mu$ với độ tin cậy 95% là

$$\theta = \mu = \left(\bar{X} \pm z_{\alpha} \sigma' \right) = \left(6.741 \pm 1.96 \times \sqrt{0.4} \right) = (5.50138715721, 7.98061284279)$$

Ta có giá trị 7 thuộc vào khoảng ước lượng, do đó chưa có cơ sở bác bỏ H_0 .

Trường hợp tiên nghiệm liên hợp

Giả sử phân phối tiên nghiệm cho tham số trung bình tổng thể $\theta = \mu$ cũng tuân theo phân phối chuẩn $N(\mu_0, \sigma_0^2)$. Khi đó, phân phối hậu nghiệm được xác định thông qua các định lý sau:

Định lý 1. Giả sử x tuân theo phân phối chuẩn $N(\theta, \sigma^2)$, với giả sử σ^2 đã biết, đồng thời phân phối tiên nghiệm của θ là $N(\mu_0, \sigma_0^2)$. Khi đó phân phối hậu nghiệm của θ là phân phối chuẩn $N(\mu_1, \sigma_1^2)$, trong đó

$$\mu_1 = \frac{\frac{x}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}}, \frac{1}{\sigma_1^2} = \frac{1}{\sigma^2} + \frac{1}{\sigma_0^2}$$

Hệ quả 1. Giả sử mẫu ngẫu nhiên x_1, x_2, \dots, x_n bao gồm n quan sát độc lập với nhau và cùng tuân theo phân phối chuẩn $N(\theta, \sigma^2)$, trong đó σ^2 đã biết và giả định phân phối tiên nghiệm cho tham số trung bình tổng thể $\theta = \mu$ là $N(\mu_0, \sigma_0^2)$. Khi đó, phân phối hậu nghiệm cho tham số trung bình tổng thể $\theta = \mu$ là $N(\mu_n, \sigma_n^2)$ trong đó

$$\mu_n = \frac{\frac{n\bar{X}}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}}{\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}}, \frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2},$$

Trong đó $\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n}$.

Ví dụ 1. Giả sử thông tin tiên nghiệm cho tham số trung bình của tổng thể là 15 đến 17. Khi đó phân phối tiên nghiệm được giả sử là $N\left(16, \frac{1}{4}\right)$, tức là $\mu_0 = 16$ và $\sigma_0 = \frac{1}{2}$.

Giả sử mẫu ngẫu nhiên gồm 10 quan sát bao gồm các giá trị 16.11, 17.37, 16.35, 15.16, 18.82, 18.12, 15.82, 16.34, 16.64, 15.0

Khi đó trung bình của mẫu là $\bar{X} = 16.573$

Độ lệch chuẩn của mẫu là $\sigma = 1.221939$

Code R:

```
data = c(16.11, 17.37, 16.35, 15.16, 18.82, 18.12, 15.82, 16.34, 16.64, 15.0)
mean(data)
sd(data)
```

Thế vào công thức trong hệ quả, ta có

$$\mu_{10} = \frac{\frac{10}{1.221939^2} \times 16.573 + \frac{16}{\frac{1}{4}}}{\frac{10}{1.221939^2} + \frac{1}{\frac{1}{4}}} = 16.3587407$$

$$\sigma_{10} = \sqrt{\frac{1}{\frac{10}{1.221939^2} + \frac{1}{\frac{1}{4}}}} = 0.3057472$$

Code R:

```
mu0 = 16
sig0 = 1/2
n = length(data)
mu10 = (n*tbx/sig^2+mu0/sig0^2)/(n/sig^2 + 1/sig0^2)
sig10 = sqrt(1/(n/sig^2 + 1/sig0^2))
c(mu10,sig10)
```

Do đó, phân phối hậu nghiệm có dạng $N(16.3587407, 0.3057472^2)$.

Kiểm định một phía về trung bình tổng thể

4. So sánh trung bình của hai mẫu ghép cặp bằng nhân tố Bayes

Bài toán kiểm định giả thuyết một phía

$$\begin{cases} H_0: \mu_d \leq 0 \\ H_1: \mu_d > 0 \end{cases}$$

Trong đó $\mu_d = \mu_1 - \mu_2$ là sự khác biệt giữa hai trung bình.

Theo thống kê Bayes, chúng ta sẽ tính $P(\mu_d \leq 0 | data)$, trong đó $data$ là các quan sát từ hai mẫu $\{x_{11}, x_{12}, \dots, x_{1n}\}$ và $\{x_{21}, x_{22}, \dots, x_{2n}\}$. Kết quả tương đương

$$P(\mu_d \leq 0 | data) = P\left(\frac{\mu_d - m'_d}{s'_d} \leq \frac{0 - m'_d}{s'_d}\right) = P\left(Z \leq -\frac{m'_d}{s'_d}\right)$$

5. So sánh trung bình của 2 mẫu độc lập

Giả định phương sai hai tổng thể đã biết

Định lý 2. Giả sử $X_1 = \{x_{11}, x_{12}, \dots, x_{1n_1}\}$ là mẫu ngẫu nhiên với cỡ mẫu n_1 chọn ra từ tổng thể thứ nhất $N(\theta_1, \sigma_1^2)$ và $X_2 = \{x_{21}, x_{22}, \dots, x_{2n_2}\}$ là mẫu ngẫu nhiên với cỡ mẫu n_2 được chọn từ tổng thể thứ hai $N(\theta_2, \sigma_2^2)$. Giả sử phân phối tiên nghiệm của θ_1 và θ_2 độc lập với nhau và cùng tuân theo phân phối đều nhận giá trị trong khoảng $(-\infty, +\infty)$. Khi đó, phân phối hậu nghiệm cho tham số $\delta = \theta_1 - \theta_2$ là phân phối chuẩn $N\left(\bar{x}_1 - \bar{x}_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$, trong đó \bar{x}_1, \bar{x}_2 tương ứng là trung bình của hai mẫu.

6. Suy luận Bayes cho tỷ lệ trên một tổng thể

Giả sử thực hiện n phép thử độc lập, trong đó kết quả cho mỗi phép thử có hai trạng thái là thành công hoặc thất bại.

Tỷ lệ số phép thử thành công trong n lần thử là θ . Khi đó, phân phối xác suất có x lần thành công trong n phép thử sẽ tuân theo phân phối nhị thức $bino(n, \theta)$. Tức là,

$$f(x|\theta) = C_n^x \theta^x (1 - \theta)^{n-x}.$$

- Nếu sử dụng tiên nghiệm đều, tức là $\pi(\theta) = 1$ thì phân phối hậu nghiệm có dạng

$$\pi(\theta|x) = C_n^x \theta^x (1 - \theta)^{n-x}.$$

Rõ ràng, trong trường hợp này, phân phối hậu nghiệm tỷ lệ với hàm hợp lý. Các kết quả ước lượng của phân phối hậu nghiệm dựa vào phân phối nhị thức.

- Nếu sử dụng tiên nghiệm liên hợp dạng phân phối beta $beta(a, b)$, tức là

$$\pi(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1}, 0 \leq \theta \leq 1.$$

Do đó, phân phối hậu nghiệm được xác định

$$\pi(\theta|x) \propto \theta^{a+x-1} (1-\theta)^{b+n-x-1}.$$

Chúng ta nhận thấy rằng, phân phối hậu nghiệm tương ứng với phân phối beta với các tham số $a' = a + x$ và $b' = b + n - y$.

Các kết quả ước lượng dựa vào phân phối beta, trong đó số yếu vị hậu nghiệm là (Bolstad, W. M., & Curran, J. M., 2016)

$$mode = \frac{a' - 1}{a' + b' - 2}$$

Trung bình hậu nghiệm (Bolstad, W. M., & Curran, J. M., 2016)

$$mean = \frac{a'}{a' + b'}$$

Phương sai hậu nghiệm

$$(s')^2 = \frac{a'b'}{(a' + b')^2(a' + b' + 1)}.$$

Kiểm định Bayes cho bài toán kiểm định giả thuyết một phía

$$\begin{cases} H_0: \theta \leq \theta_0 \\ H_1: \theta > \theta_0 \end{cases}$$

Với mức ý nghĩa α .

Tính xác suất hậu nghiệm của giả thuyết

$$P(H_0: \theta \leq \theta_0|x) = \int_0^{\theta_0} \pi(\theta|x) d\theta.$$

Bác bỏ H_0 nếu xác suất hậu nghiệm nhỏ hơn mức ý nghĩa α .

Bài toán kiểm định hai phía

$$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$$

Tính khoảng ước lượng $(1 - \alpha) \times 100\%$ cho tham số θ , nếu có chứa giá trị θ_0 thì chấp nhận H_0 , ngược lại bác bỏ H_0 .

Code in R

```
pbeta( $\theta_0$ , shape1 = 1, shape2 = 5, lower.tail = TRUE, log.p = FALSE)
```