TRƯỜNG HÈ
THỐNG KÊ BAYES VÀ ỨNG DỤNG
Mở đầu về Thống kê Bayes

Ngô Hoàng Long (Trường ĐH Sư phạm Hà Nội)
Trần Minh Ngọc (Đại học Sydney, Úc)

HỘI NGHỊ TOÀN QUỐC LẦN THỨ VII XÁC SUẤT-THÔNG KÊ
Quy Nhơn 4/8/2025

Câu hỏi

- Thầy Sơn có một đồng xu.
- Xác suất xuất hiện mặt sấp trong mỗi lần gieo đều bằng $1/2$.
- Thầy Sơn gieo đồng xu 10 lần đều thấy xuất hiện mặt sấp, tính xác suất lần gieo thứ 11 cũng xuất hiện mặt sấp.

Công thức Bayes

- $\{\Omega, \mathcal{F}, \mathbb{P}\}$ là một không gian xác suất.
- Xét (A_k) là một phân hoạch của Ω trong $\mathcal{F}$.
- $B \in \mathcal{F}$ là một biến cố bất kì.

Khi đó

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_k \mathbb{P}(B|A_k)\mathbb{P}(A_k)}.$$

Ta gọi

- $\mathbb{P}(A_i)$ là *xác suất tiên nghiệm* của A_i.
- $\mathbb{P}(A_i|B)$ là *xác suất hậu nghiệm* của A_i.

Ví dụ 1

- Thầy Ngọc có 3 đồng xu có vẻ ngoài giống nhau, xác suất xuất hiện mặt sấp trong mỗi lần gieo của 3 đồng xu trên lần lượt là $\frac{1}{4}, \frac{1}{2}$ và $\frac{3}{4}$.
- Thầy Ngọc chọn ngẫu nhiên 1 trong 3 đồng xu trên. Gọi p là xác suất xuất hiện mặt sấp của đồng xu được chọn.
- p là một đại lượng ngẫu nhiên nhận các giá trị: $\frac{1}{4}, \frac{1}{2}$ và $\frac{3}{4}$ với cùng xác suất là $\frac{1}{3}$.
- Xét các biến cố $A_1 = \left\{p = \frac{1}{4}\right\}, A_2 = \left\{p = \frac{1}{2}\right\}$ và $A_3 = \left\{p = \frac{3}{4}\right\}$. Ta có $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \mathbb{P}(A_3) = \frac{1}{3}$.
- Gọi S là biến cố khi thầy Ngọc gieo đồng xu thu được mặt sấp. Ta có:

$$\mathbb{P}(S|A_1) = \frac{1}{4},\ \mathbb{P}(S|A_2) = \frac{1}{2} \text{ và } \mathbb{P}(S|A_3) = \frac{3}{4}.$$

Ví dụ 1

- Theo công thức xác suất toàn phần :

$$\mathbb{P}(S) = \sum_i \mathbb{P}(S|A_i)\mathbb{P}(A_i) = \frac{1}{2}.$$

- Câu hỏi: Biết lần đầu thầy Ngọc gieo được mặt sấp, tính xác suất lần thứ hai thầy Ngọc cũng gieo được mặt sấp?
- Trả lời: Do hai lần gieo là độc lập nên xác suất lần thứ hai gieo được mặt sấp là $\frac{1}{2}$ (không phụ thuộc vào kết quả lần gieo thứ nhất)???

Ví dụ 1

- Theo công thức xác suất toàn phần :

$$\mathbb{P}(S) = \sum_i \mathbb{P}(S|A_i)\mathbb{P}(A_i) = \frac{1}{2}.$$

- Câu hỏi: Biết lần đầu thầy Ngọc gieo được mặt sấp, tính xác suất lần thứ hai thầy Ngọc cũng gieo được mặt sấp?
- Trả lời: Do hai lần gieo là độc lập nên xác suất lần thứ hai gieo được mặt sấp là $\frac{1}{2}$ (không phụ thuộc vào kết quả lần gieo thứ nhất)???
- Thực tế: $\mathbb{P}(S_2|S_1) = \frac{7}{12} > \frac{1}{2}$.

Ví dụ 1

- Theo công thức xác suất toàn phần :

$$\mathbb{P}(S) = \sum_i \mathbb{P}(S|A_i)\mathbb{P}(A_i) = \frac{1}{2}.$$

- Câu hỏi: Biết lần đầu thầy Ngọc gieo được mặt sấp, tính xác suất lần thứ hai thầy Ngọc cũng gieo được mặt sấp?
- Trả lời: Do hai lần gieo là độc lập nên xác suất lần thứ hai gieo được mặt sấp là $\frac{1}{2}$ (không phụ thuộc vào kết quả lần gieo thứ nhất)???
- Thực tế: $\mathbb{P}(S_2|S_1) = \frac{7}{12} > \frac{1}{2}$.
- Như vậy, kết quả các lần gieo là *không độc lập* với nhau.

Ví dụ 1

- Theo công thức xác suất toàn phần :

$$\mathbb{P}(S) = \sum_i \mathbb{P}(S|A_i)\mathbb{P}(A_i) = \frac{1}{2}.$$

- Câu hỏi: Biết lần đầu thầy Ngọc gieo được mặt sấp, tính xác suất lần thứ hai thầy Ngọc cũng gieo được mặt sấp?
- Trả lời: Do hai lần gieo là độc lập nên xác suất lần thứ hai gieo được mặt sấp là $\frac{1}{2}$ (không phụ thuộc vào kết quả lần gieo thứ nhất)???
- Thực tế: $\mathbb{P}(S_2|S_1) = \frac{7}{12} > \frac{1}{2}$.
- Như vậy, kết quả các lần gieo là *không độc lập* với nhau.

Ví dụ 1

- Nếu thầy Ngọc gieo được mặt sấp thì xác suất của biến cố A_1 là:

$$\mathbb{P}(A_1|S) = \frac{\mathbb{P}(S|A_1)\mathbb{P}(A_1)}{\mathbb{P}(S)} = \frac{\frac{1}{4}.\frac{1}{3}}{\frac{1}{2}} = \frac{1}{6}.$$

- Tương tự ta có

$$\mathbb{P}(A_2|S) = \frac{1}{3} \quad , \quad \mathbb{P}(A_3|S) = \frac{1}{2}.$$

Như vậy, nếu thầy Ngọc gieo được mặt sấp thì khả năng đồng xu được chọn có $p = \frac{3}{4}$ là cao nhất.

Ví dụ 1

- Nếu thầy Ngọc gieo được mặt sấp thì xác suất của biến cố A_1 là:

$$\mathbb{P}(A_1|S) = \frac{\mathbb{P}(S|A_1)\mathbb{P}(A_1)}{\mathbb{P}(S)} = \frac{\frac{1}{4}.\frac{1}{3}}{\frac{1}{2}} = \frac{1}{6}.$$

- Tương tự ta có

$$\mathbb{P}(A_2|S) = \frac{1}{3} \quad , \quad \mathbb{P}(A_3|S) = \frac{1}{2}.$$

Như vậy, nếu thầy Ngọc gieo được mặt sấp thì khả năng đồng xu được chọn có $p = \frac{3}{4}$ là cao nhất.

Ví dụ 1

Giả sử thầy Ngọc gieo 4 lần liên tiếp đều được mặt sấp thì xác suất của biến cố A_1 là:

$$\begin{aligned}\mathbb{P}(A_1|SSSS) &= \frac{\mathbb{P}(SSSS|A_1)\mathbb{P}(A_1)}{\mathbb{P}(SSSS|A_1)\mathbb{P}(A_1) + \mathbb{P}(SSSS|A_2)\mathbb{P}(A_2) + \mathbb{P}(SSSS|A_3)\mathbb{P}(A_3)} \\ &= \frac{1}{98}.\end{aligned}$$

Tương tự ta có

$$\mathbb{P}(A_2|SSSS) = \frac{16}{98} \quad , \quad \mathbb{P}(A_3|SSSS) = \frac{81}{98}.$$

Ví dụ 2

- Bạn Hà có một đồng xu, xác suất gieo đồng xu được mặt sấp là p.
- Giả sử p là đại lượng ngẫu nhiên có phân phối đều $U[0,1]$.
- Như trong ví dụ 1, ta sẽ phân tích sự thay đổi phân phối của p dựa trên kết quả các lần gieo.
- Giả sử khi gieo đồng xu xuất hiện mặt sấp, khi đó xác suất để $p < x$ với $x \in [0,1]$ là

$$\mathbb{P}[p < x|S] = \frac{\mathbb{P}[p < x, S]}{\mathbb{P}(S)}, \quad \mathbb{P}[p < x] = x \quad \forall x \in [0,1].$$

Với $x \in (0,1), n > 0$,

$$\begin{aligned}
\mathbb{P}(S \cap \{p < x\}) &= \sum_{k=0}^{n-1} \mathbb{P}\left(S \cap \left\{\frac{kx}{n} \leq p < \frac{(k+1)x}{n}\right\}\right) \\
&= \sum_{k=0}^{n-1} \mathbb{P}\left(S \middle| \frac{kx}{n} \leq p < \frac{(k+1)x}{n}\right) \mathbb{P}\left(\frac{kx}{n} \leq p < \frac{(k+1)x}{n}\right) \\
&\approx \sum_{k=0}^{n-1} \frac{kx}{n}\frac{x}{n} \approx \frac{x^2}{2}.
\end{aligned}$$

Do đó, $\mathbb{P}[p < x|S] = x^2$.
Vậy hàm mật độ của p với điều kiện S là :

$$f_{p|S}(x) = \frac{\partial}{\partial x}\mathbb{P}[p < x|S] = \begin{cases} 2x & \text{nếu } x \in [0,1] \\ 0 & \text{nếu } x \neq [0,1]. \end{cases}$$

Tương tự, hàm mật độ của p với điều kiện N là:

$$f_{p|N}(x) = \begin{cases} 2 - 2x & \text{nếu } x \in [0,1] \\ 0 & \text{nếu } x \neq [0,1]. \end{cases}$$

Công thức Bayes tổng quát

- Giả sử $\mathbb{X} = (X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên có hàm mật độ $f(x|\theta)$ trong đó θ là tham số chưa biết.
- Ta cũng giả sử θ là một đại lượng ngẫu nhiên có hàm mật độ $p(\theta)$. Hàm mật độ $p(\theta)$ được gọi là *tiên nghiệm*.
- Dựa trên kết quả quan sát $X_1, X_2, ..., X_n$, ta xác định lại hàm mật độ của θ, tức là $p(\theta|X_1, .., X_n)$. Hàm mật độ $p(\theta|X_1, .., X_n)$ được gọi là *hậu nghiệm* của θ.

Công thức Bayes cho ta:

$$p(\theta|\mathbb{X}) = \frac{f(\mathbb{X}|\theta)p(\theta)}{\int f(\mathbb{X}|\theta')p(\theta')d\theta'}.$$

$$p(\theta|\mathbb{X}) = \frac{f(\mathbb{X}|\theta)p(\theta)}{\int f(\mathbb{X}|\theta')p(\theta')d\theta'}.$$

Lưu ý:

- $p(\theta|\mathbb{X})$ là một hàm của θ.
- $\int f(\mathbb{X}|\theta')p(\theta')d\theta'$ không phụ thuộc vào θ.
  Khi chỉ quan tâm đến θ thì ta viết $(*)$ dưới dạng

$$p(\theta|\mathbb{X}) \propto f(\mathbb{X}|\theta)p(\theta).$$

Hậu nghiệm $\propto$ tiên nghiệm $\times$ hàm hợp lý

Bài tập

Áp dụng công thức Bayes để tính hàm mật độ điều kiện trong Ví dụ 2.

Suy luận Bayes

- θ : véc tơ tham số trong không gian Θ.
- $\boldsymbol{y} = \{y_1, \ldots, y_n\}$: dữ liệu
- Tiên nghiệm (Prior distribution): $p(\theta)$
- Hàm hợp lý (Likelihood): $p(y \mid \theta)$ Suy luận Bayes dựa trên phân phối hậu nghiệm với mật độ

$$p(\theta \mid \boldsymbol{y}) = \frac{p(\theta)p(\boldsymbol{y} \mid \theta)}{p(\boldsymbol{y})}$$

trong đó

$$p(y) = \int_{\Omega} p(\theta)p(y \mid \theta)d\theta$$

là hằng số chuẩn hoá, thường được gọi là hàm hợp lý biên duyên (marginal likelihood).

Chú ý

- Phân phối tiên nghiệm $p(\theta)$ được xác định dựa trên thông tin/kiến thức tiên nghiệm, hoặc đơn giản là vì lý do toán học, hay vì sự tiện lợi.
- Phân phối tiên nghiệm được gọi là đứng đấn (proper) nếu $\int p(\theta)d\theta = 1$, ngược lại thì gọi là không đứng đắn (improper).
- Hàm hợp lý $p(y \mid \theta)$ được xác định dựa trên mô hình xác suất.
- Hằng số chuẩn hóa $p(y)$ thường không biết được, do đó ta thường viết:

$$p(\theta \mid y) \propto p(\theta)p(y \mid \theta).$$

- Trong hầu hết các trường hợp, phân phối hậu nghiệm $p(\theta \mid y)$ là không tính được nên các phương pháp Monte Carlo thường được sử dụng để nghiên cứu $p(\theta \mid y)$.

Người ta thường dùng mean và mode là các số đặc trưng cho xu thế trung tâm của hậu nghiệmm. Mean hậu nghiệm là

$$\mathbb{E}(\theta \mid \boldsymbol{y}) = \int \theta p(\theta \mid \boldsymbol{y}) d\theta$$

Đề đặc trưng cho độ phân tấn của hậu nghiệm, ta có thể dùng độ lệch chuẩn (hoặc khoảng tứ phân vị)...

$$\mathbb{V}(\theta \mid \boldsymbol{y}) = \int (\theta - \mu)^2 p(\theta \mid \boldsymbol{y}) d\theta$$

with $\mu = \mathbb{E}(\theta \mid \boldsymbol{y})$.

Ước lượng hậu nghiệm cực đại (Maximum a posteriori estimator)

Ước lượng hậu nghiệm cực đại (MAP) là đại lượng

$$\widehat{\theta}_{\mathrm{MAP}} = \operatorname*{argmax}_{\theta}\{\log(p(y \mid \theta)) + \log(p(\theta))\}$$

Khoảng ước lượng Bayes

- Giả sử $\alpha \in (0, 1)$.
- Giả sử a và b là các hằng số thoả mãn

$$\int_{-\infty}^{a} p(\theta \mid y) d\theta = \int_{b}^{\infty} p(\theta \mid y) d\theta = \alpha/2$$

- Đặt $C = (a, b)$.

Khi đó

$$\mathbb{P}(\theta \in C \mid \boldsymbol{y}) = \int_{a}^{b} p(\theta \mid \boldsymbol{y}) d\theta = 1 - \alpha$$

C được gọi là khoảng tin cậy/khoảng xác tín (credible interval) hậu nghiệm $1 - \alpha$.

Phân phối dự đoán (Predictive distribution)

- Giả sử y^* là một điểm dữ liệu trong tương lai.
- $\mathbf{y} = \{y_1, \ldots, y_n\}$ là tập dữ liệu đã quan sát.
- Ta muốn dự đoán y^*.
- Khi chưa có dữ liệu nào, phân phối của y^* là:

$$p(y^*) = \int p(\theta) p(y^* \mid \theta)\, d\theta$$

- Khi đã có các quan sát $\boldsymbol{y}$ ta thay tiên nghiệm $p(\theta)$ bởi hậu nghiệm $p(\theta \mid \boldsymbol{y})$, phân phối của y^* trở thành

$$p\left(y^* \mid \boldsymbol{y}\right) = \int p(\theta \mid \boldsymbol{y}) p\left(y^* \mid \theta\right) d\theta$$

- $p\left(y^* \mid \boldsymbol{y}\right)$ được gọi là phân phối dự đoán (hậu nghiệm). Ta sử dụng $p\left(y^* \mid \boldsymbol{y}\right)$ để suy diễn y^*.
- Ước lượng điểm của y^* là

$$\mathbb{E}\left(y^* \mid y\right) = \int y^* p\left(y^* \mid y\right) dy^*$$

và sai số của ước lượng được đặc trưng bởi

$$\mathbb{V}\left(y^* \mid \boldsymbol{y}\right) = \int \left(y^* - \mathbb{E}\left(y^* \mid \boldsymbol{y}\right)\right)^2 p\left(y^* \mid \boldsymbol{y}\right) dy^*.$$

Ví dụ 3

- Gọi θ là tỉ lệ trẻ sơ sinh nữ.
- Gọi $\boldsymbol{y} = \{y_1, \ldots, y_n\}$ là mẫu số liệu về giới tính của n trẻ sơ sinh, trong đó $y_i = 1$ nếu trẻ thứ i là nữ.
- Giả sử y_i có phân phối nhị thức $\text{Bi}(1, \theta)$: $p\left(y_i \mid \theta\right) = \theta^{y_i}(1-\theta)^{1-y_i}$
- Hàm hợp lý

$$p(\boldsymbol{y} \mid \theta) = \prod_{i=1}^{n} p\left(y_i \mid \theta\right) = \theta^k(1-\theta)^{n-k}, \quad k = \sum y_i$$

Phân phối beta

Đại lượng ngẫu nhiên θ được gọi là có phân phối $Beta(\alpha, \beta)$; $\alpha, \beta > 0$, nếu nó nhận giá trị trên đoạn $[0, 1]$ với hàm mật độ :

$$p(\theta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} & \text{nếu } \theta \in [0,1] \\ 0 & \text{nếu } \theta \notin [0,1]. \end{cases}$$

Tính chất:

- $E[\theta] = \dfrac{\alpha}{\alpha+\beta}$.
- $Var(\theta) = \dfrac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$.
- $Mode\theta = \dfrac{\alpha-1}{\alpha-1+\beta-1}$ nếu $\alpha > 1; \beta > 1$.

Ví dụ 3

- Giả sử θ có tiên nghiệm là phân phối đều trên đoạn $[0, 1]$.
- Phân phối hậu nghiệm của θ là

$$p(\theta \mid y) \propto \theta^k (1-\theta)^{n-k}$$

- Phân phối dự đoán

$$\begin{aligned} P(y^* = 1 \mid y) &= \int P(y^* = 1 \mid \theta)\, p(\theta \mid y) d\theta \\ &= \int \theta p(\theta \mid y) d\theta \\ &= \frac{k+1}{n+2}. \end{aligned}$$

Ví dụ 4

- Giả sử $y_i \sim \mathcal{N}\left(\mu, \sigma^2\right), i = 1, \ldots, n; \sigma^2$ là đã biết.
- Giả sử μ có tiên nghiệm là $p(\mu) \sim \mathcal{N}\left(\mu_0, \sigma_0^2\right)$.
- Hậu nghiệm của μ xác định bởi

$$\begin{aligned} p(\mu \mid \boldsymbol{y}) &\propto p(\mu) p(\boldsymbol{y} \mid \mu) \\ &\propto \exp\left(-\frac{1}{2\sigma_0^2}\left(\mu - \mu_0\right)^2\right) \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \mu\right)^2\right) \\ &\propto \exp\left(-\frac{1}{2\sigma_*^2}\left(\mu - \mu_*\right)^2\right), \end{aligned}$$

trong đó

$$\mu_* = \frac{\bar{y} + \frac{\sigma^2 \mu_0}{n\sigma_0^2}}{1 + \frac{\sigma^2}{n\sigma_0^2}} = w\bar{y} + (1 - w)\mu_0, \quad w = \frac{1}{1 + \frac{\sigma^2}{n\sigma_0^2}} \in (0, 1)$$

$$\sigma_*^2 = \frac{\sigma^2}{n + \sigma^2/\sigma_0^2}$$

Do đó $p(\mu \mid \boldsymbol{y}) \sim \mathcal{N}\left(\mu_*, \sigma_*^2\right)$.

Ví dụ 4

- $\widehat{\mu} = \mu_*$ thường được dùng làm ước lượng điểm cho μ.
- Một 95% khoảng xác tín (credible interval) cho μ là $\mu_* \pm 1.96\sigma_*$.

> Xác suất μ thuộc khoảng cố định trên là 0.95.

> Khoảng xác tín là cố định (khi dữ liệu đã được thu thập), μ là ngẫu nhiên.

Đây là điểm khác biệt cơ bản với cách diễn đạt khoảng tin cậy theo nghĩa tần số

$$\mu_* = w\bar{y} + (1-w)\mu_0, \quad w = \frac{1}{1+\frac{\sigma^2}{n\sigma_0^2}} \in (0,1)$$

μ_* là trung bình có trọng số của ước lượng tần số $\bar{y}$ và trung bình tiên nghiệm μ_0.

- $\mu_* \to \bar{y}$ khi $n \to \infty$: Ước lượng Bayes hội tụ đến ước lượng tần số khi cỡ mẫu đủ lớn, tức là thông tin từ dữ liệu áp đảo thông tin tiên nghiệm.
- $\mu_* \to \bar{y}$ khi $\sigma_0^2 \to \infty$: Ước lượng Bayes hội tụ tới ước lượng tần số khi tiên nghiệm không chứa nhiều thông tin.

Tiên nghiệm liên hợp

- Nếu phân phối tiên nghiệm và phân phối hậu nghiệm thuộc cùng một họ phân phối, ta nói rằng phân phối tiên nghiệm là liên hợp (conjugate) đối với mô hình.
- Ví dụ: trong ví dụ trước, phân phối tiên nghiệm chuẩn (normal) là liên hợp với mô hình dữ liệu cũng là phân phối chuẩn. Điều này đúng vì cả phân phối tiên nghiệm và hậu nghiệm của μ đều là các phân phối chuẩn (nhưng với các tham số khác nhau).
- Việc sử dụng phân phối tiên nghiệm liên hợp giúp ta dễ dàng xác định phân phối hậu nghiệm.
- Tuy nhiên, không phải lúc nào cũng có thể sử dụng được phân phối tiên nghiệm liên hợp.

Mô hình Bernoulli / Nhị thức (Binomial)

- Dữ liệu: $y \sim \text{Binomial}(n, \theta)$ hoặc $y_i \sim \text{Bernoulli}(\theta)$.
- Tiên nghiệm liên hợp: $\theta \sim \text{Beta}(\alpha, \beta)$.
- Hậu nghiệm: $\theta \mid y \sim \text{Beta}(\alpha + y, \beta + n - y)$.

Mô hình Poisson

- Dữ liệu: $y \sim \text{Poisson}(\lambda)$.
- Tiên nghiệm liên hợp: $\lambda \sim \text{Gamma}(\alpha, \beta)$.
- Hậu nghiệm: $\lambda \mid y \sim \text{Gamma}(\alpha + y, \beta + 1)$.

Mô hình Gaussian (trung bình biết - phương sai chưa biết)

- Dữ liệu: $y_i \sim \mathcal{N}\left(\mu, \sigma^2\right)$, với μ biết và σ^2 chưa biết.
- Tiên nghiệm liên hợp: $\sigma^2 \sim$ Inverse-Gamma(α, β).
- Hậu nghiệm: $\sigma^2 \mid y \sim$ Inverse-Gamma (α', β').

Mô hình Gaussian (trung bình chưa biết - phương sai biết)

- Dữ liệu: $y_i \sim \mathcal{N}\left(\mu, \sigma^2\right)$, với σ^2 đã biết.
- Tiên nghiệm liên hợp: $\mu \sim \mathcal{N}\left(\mu_0, \tau^2\right)$.
- Hậu nghiệm: $\mu \mid y \sim \mathcal{N}\left(\mu_n, \tau_n^2\right)$.

Mô hình Gaussian (cả trung bình và phương sai chưa biết)

- Tiên nghiệm liên hợp: phân phối Normal-Inverse-Gamma

$$\left(\mu, \sigma^2\right) \sim \text{ Normal-Inverse-Gamma } \left(\mu_0, \lambda, \alpha, \beta\right).$$

- Hậu nghiệm cũng thuộc cùng họ phân phối.