

1 Phần xác suất

1.1 Các công thức xác suất

Công thức cộng và nhân xác suất:

- $P(A + B) = P(A) + P(B) - P(AB)$, và $P(\sum_{i=1}^n A_i) = \sum_i P(A_i) - \sum_{i < j} P(A_i A_j) + \sum_{i < j < k} P(A_i A_j A_k) - \dots$
- $P(AB) = P(A)P(B|A)$ và $P(A_1 A_2 \dots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \dots P(A_n|A_1 A_2 \dots A_{n-1})$

Với A_1, \dots, A_n là một họ các biến cố đầy đủ:

- Công thức xác suất đầy đủ: $P(F) = P(A_1)P(F|A_1) + P(A_2)P(F|A_2) + \dots + P(A_n)P(F|A_n)$.
- Công thức Bayes: $P(A_k|F) = \frac{P(A_k)P(F|A_k)}{P(F)}$.

1.2 Biến ngẫu nhiên (BNN)

- BNN X rời rạc: $E(X) = \sum_i x_i p_i$, và $D(X) = \sum_i (x_i - E(X))^2 p_i = \sum_i x_i^2 p_i - [E(X)]^2$.
- BNN X liên tục: $E(X) = \int_{-\infty}^{\infty} x f(x) dx$, và $D(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - [E(X)]^2$.

1.3 Các hàm phân phối xác suất cơ bản

Phân phối nhị thức, $X \sim B(n, p)$: $P(X = k) = C_n^k p^k q^{n-k}$, $k = 1, \dots, n$, và $E(X) = np$, $D(X) = npq$.

Phân phối siêu bội, $X \sim H(N, M, n)$: $P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$, và $E(X) = np$, $D(X) = np(1-p) \left(\frac{N-n}{N-1} \right)$, $p = \frac{M}{N}$.

Phân phối Poisson, $X \sim P(a)$: $P(X = k) = \frac{e^{-a} a^k}{k!}$, $k = 1, 2, \dots$ và $E(X) = D(X) = a$.

Phân phối mũ, $X \sim E(\lambda)$: $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ và $E(X) = \frac{1}{\lambda}$, $D(X) = \frac{1}{\lambda^2}$.

Phân phối chuẩn, $X \sim N(a, \sigma^2)$: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$ và $E(X) = a$, $D(X) = \sigma^2$.

Định lý giới hạn trung tâm: Nếu X_1, \dots, X_n là đôi một độc lập và $E(X_k) = a$, $D(X_k) = \sigma^2$, $\bar{X} = \frac{\sum_{k=1}^n X_k}{n}$, khi n đủ lớn thì $\frac{\bar{X} - a}{\sigma/\sqrt{n}} \sim N(0, 1)$.

2 Phần thống kê

2.1 Khoảng tin cậy

Khoảng tin cậy cho kỳ vọng :

- Biết σ^2 , X có phân phối chuẩn hoặc n đủ lớn: $\bar{x} - Z_\alpha \frac{\sigma}{\sqrt{n}} \leq a \leq \bar{x} + Z_\alpha \frac{\sigma}{\sqrt{n}}$
- Không biết σ^2 , cỡ mẫu $n \geq 30$: $\bar{x} - Z_\alpha \frac{s}{\sqrt{n}} \leq a \leq \bar{x} + Z_\alpha \frac{s}{\sqrt{n}}$
- X có phân phối chuẩn, không biết σ^2 , cỡ mẫu $n < 30$: $\bar{x} - t_{\alpha/2}^{n-1} \frac{s}{\sqrt{n}} \leq a \leq \bar{x} + t_{\alpha/2}^{n-1} \frac{s}{\sqrt{n}}$

Khoảng tin cậy cho tỷ lệ tổng thể P , $n > 30$: $f = \frac{m}{n}$, $f - Z_\alpha \sqrt{\frac{f(1-f)}{n}} \leq P \leq f + Z_\alpha \sqrt{\frac{f(1-f)}{n}}$.

2.2 Kiểm định giả thuyết thống kê, một mẫu

Kiểm định cho kỳ vọng $H : a = a_0$ vs $\bar{H} : a \neq a_0$:

1. Biết σ^2 , X có phân phối chuẩn hoặc n đủ lớn: $U_{qs} = \frac{\bar{X} - a_0}{\sigma/\sqrt{n}} \sim N(0, 1)$. Nếu $|U_{qs}| > Z_\alpha$: Bác bỏ H .
2. $n \geq 30$, không biết σ^2 : $U_{qs} = \frac{\bar{X} - a_0}{s/\sqrt{n}} \sim N(0, 1)$. Nếu $|U_{qs}| > Z_\alpha$: Bác bỏ H .
3. $n < 30$, X có phân phối chuẩn, không biết σ^2 : $T_{qs} = \frac{\bar{X} - a_0}{s/\sqrt{n}} \sim student(n-1)$. Nếu $|T_{qs}| > t_{\alpha/2}^{n-1}$: Bác bỏ H .

Kiểm định cho tỉ lệ tổng thể P , $n > 30$: $H : P = P_0$ vs $\bar{H} : P \neq P_0$: $U_{qs} = \frac{f - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} \sim N(0, 1)$. Nếu $|U_{qs}| > Z_\alpha$: Bác bỏ H .

Bảng tiêu chuẩn bác bỏ H_0 mở rộng :

GT H_0	GT đối H_1	Miền bác bỏ khi dùng KD $N(0, 1)$	Miền bác bỏ khi dùng KD $T(n-1)$
$a = a_0$ hoặc $p = p_0$	$a \neq a_0$ hoặc $p \neq p_0$	$W_\alpha = (-\infty, -Z_\alpha) \cup (Z_\alpha, \infty)$	$W_\alpha = (-\infty, -t_{\alpha/2}^{n-1}) \cup (t_{\alpha/2}^{n-1}, \infty)$
	$a < a_0$ hoặc $p < p_0$	$W_\alpha = (-\infty, -Z_{2\alpha})$	$W_\alpha = (-\infty, -t_{\alpha}^{n-1})$
	$a > a_0$ hoặc $p > p_0$	$W_\alpha = (Z_{2\alpha}, \infty)$	$W_\alpha = (t_{\alpha}^{n-1}, \infty)$

2.3 Kiểm định giả thuyết thống kê, hai mẫu

Kiểm định cho kỳ vọng : Giả thuyết $H : a_1 = a_2$ vs $\bar{H} : a_1 \neq a_2$

1. Đã biết phương sai, phân phối chuẩn: $U_{qs} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$. Nếu $|U_{qs}| > Z_\alpha$: Bác bỏ H .
2. Chưa biết phương sai, và $n_1, n_2 > 30$: $U_{qs} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim N(0, 1)$. Nếu $|U_{qs}| > Z_\alpha$: Bác bỏ H .
3. Chưa biết phương sai, phân phối chuẩn $n_1 < 30$ hoặc $n_2 < 30$: $T_{qs} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim student(n_1 + n_2 - 2)$. Nếu $|T_{qs}| > t_{\alpha/2}^{n_1+n_2-2}$:
Bác bỏ H .

Kiểm định cho tỉ lệ tổng thể, $n_1, n_2 > 30$: $H : P_1 = P_2$ vs $\bar{H} : P_1 \neq P_2$: $U_{qs} = \frac{f_1 - f_2}{\sqrt{p_0(1-p_0)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$, với
 $p_0 = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$. Nếu $|U_{qs}| > Z_\alpha$: Bác bỏ H .

Các trường hợp kiểm định một phía làm tương tự như trường hợp một mẫu.

2.4 Phân tích phương sai (ANOVA) một nhân tố, cỡ mẫu bằng nhau

Quan sát một mẫu có $N = kn$ giá trị quan trắc, trong đó k là số phương thức xử lý của nhân tố, và mỗi phương thức xử lý có n giá trị quan trắc.

Bài toán kiểm định: $H_0 : \tau_1 = \tau_2 = \dots = \tau_k = 0$ vs $H_1 : \tau_i \neq 0$, với ít nhất một i . Bác bỏ H_0 khi: $F = \frac{MSB}{MSW} > F_{\alpha; k-1, k(n-1)}$.

Nguồn của sự biến thiên	SS	df	MS	F
Giữa các nhóm(SSB)	$SSB = n \sum_{i=1}^k (\bar{y}_i - \bar{y}..)^2 = \sum_{i=1}^k \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{N}$	$k - 1$	$MSB = \frac{SSB}{k-1}$	$F = \frac{MSB}{MSW}$
Trong từng nhóm (SSW)	$SSW = \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}_i.)^2 = SST - SSB$	$k(n - 1)$	$MSW = \frac{SSW}{k(n-1)}$	
Tổng (SST)	$SST = \sum_{i=1}^k \sum_{j=1}^n (y_{ij} - \bar{y}..)^2 = \sum_{i=1}^k \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{N}$	$kn - 1$		

2.5 Hồi quy tuyến tính đơn

Mô hình hồi quy tuyến tính mẫu Y theo $X : y = A + Bx$ với $B = \frac{\bar{xy} - \bar{x}\bar{y}}{\hat{s}_X^2}$ và $A = \bar{y} - B\bar{x}$.

Hệ số tương quan mẫu: $r_{xy} = \frac{\bar{xy} - \bar{x}\bar{y}}{\hat{s}_X \hat{s}_Y}$.