Đại Học Bách Khoa TP.Hồ Chí Minh
Bộ môn Toán Ứng Dụng

TÓM TẮT MỘT SỐ CÔNG THỨC
Môn: XÁC SUẤT THỐNG KÊ

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 $P\left(\sum_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) - \dots$
- $P(AB) = P(A)P(B|A)$ và $P(A_1A_2...A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)...P(A_n|A_1A_2...A_{n-1})$

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 Bayse: $P(A_k|F) = \dfrac{P(A_k)P(F|A_k)}{P(F)}$.

1.2 Biến ngẫu nhiên (BNN):

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

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 $\mathbb{E}(X) = np$, $\mathbb{V}ar(X) = npq$.

Phân phối Poisson, $X \sim P(\lambda)$ $P(X=k) = \dfrac{e^{-\lambda}\lambda^k}{k!},\ k = 1, 2, \dots,$ và $\mathbb{E}(X) = \mathbb{V}ar(X) = \lambda$.

Phân phối siêu bội, $X \sim H(N,K,n)$: $P(X=k) = \dfrac{C_K^k\, C_{N-K}^{n-k}}{C_N^n}$ và $\mathbb{E}(X) = np$, $\mathbb{V}ar(X) = np(1-p)\left(\dfrac{N-n}{N-1}\right)$, $p = \dfrac{K}{N}$.

Phân phối mũ, $X \sim Exp(\lambda)$: $f(x) = \begin{cases} \lambda e^{-\lambda x},\ x \geq 0 \\ 0,\ x < 0 \end{cases}$, và $\mathbb{E}(X) = \dfrac{1}{\lambda}$, $\mathbb{V}ar(X) = \dfrac{1}{\lambda^2}$.

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

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

2 Phần thống kê

2.1 Khoảng tin cậy

Khoảng tin cậy cho kỳ vọng :

- Biết σ^2, X có phân phối chuẩn hoặc cỡ mẫu n đủ lớn: $\overline{x} - z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
- Không biết σ^2, và X có phân phối chuẩn: $\overline{x} - t_{\alpha/2}^{n-1}\dfrac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + t_{\alpha/2}^{n-1}\dfrac{s}{\sqrt{n}}$

Khoảng tin cậy cho tỷ lệ tổng thể P, $n > 30$: $\hat{P} - z_{\alpha/2}\sqrt{\dfrac{\hat{P}(1-\hat{P})}{n}} \leq P \leq \hat{P} + z_{\alpha/2}\sqrt{\dfrac{\hat{P}(1-\hat{P})}{n}}$. Trong đó: $\hat{P} = \dfrac{X}{n}$, X là số phần tử thoả tính chất $\mathcal{A}$ trong mẫu gồm n phần tử.

2.2 Kiểm định giả thuyết thống kê, một mẫu

Kiểm định cho kỳ vọng :

1. Biết σ^2, X có phân phối chuẩn hoặc cỡ mẫu n đủ lớn: $z_0 = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$ ==> Dùng bảng 1.
2. Không biết σ^2 và X có phân phối chuẩn: $t_0 = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$ ==> Dùng bảng 2.
3. Không biết σ^2, X có phân phối bất kỳ, cỡ mẫu đủ lớn: $z_0 = \dfrac{\overline{X} - \mu_0}{s/\sqrt{n}}$ ==> Dùng bảng 1.

Kiểm định cho tỉ lệ tổng thể, $n > 30$: $z_0 = \dfrac{\hat{P} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$ ==> Dùng bảng 1. Trong đó: $\hat{P} = \dfrac{X}{n}$, X là số phần tử thoả tính chất $\mathcal{A}$ trong mẫu gồm n phần tử.

2.3 Kiểm định giả thuyết thống kê, hai mẫu

Kiểm định cho kỳ vọng :

1. Biết phương sai, phân phối chuẩn hoặc cỡ mẫu đủ lớn: $z_0 = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ ==> Dùng bảng 1.
2. Chưa biết phương sai, có phân phối chuẩn và cỡ mẫu đủ lớn: $z_0 = \dfrac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ ==> Dùng bảng 1.
3. Chưa biết phương sai, có phân phối chuẩn, cỡ mẫu nhỏ và $\sigma_1 = \sigma_2$:
$S_p = \dfrac{(n_-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$, $t_0 = \dfrac{\overline{X} - \overline{Y}}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ ==> Dùng bảng 2 với $df = n_1 + n_2 - 2$.

Kiểm định cho tỉ lệ tổng thể, $n_1, n_2 > 30$: $z_0 = \dfrac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$ ==> Dùng bảng 1. Trong đó: $\hat{P} = \dfrac{X+Y}{n_1+n_2}$, $\hat{P}_1 = \dfrac{X}{n_1}$, $\hat{P}_2 = \dfrac{Y}{n_2}$, X và Y lần lượt là số phần tử thoả tính chất $\mathcal{A}$ trong mẫu gồm n_1 và n_2 phần tử.

Bảng quy tắc bác bỏ H_0:

Đối thuyết H1	Miền bác bỏ	Trị số p_v
Hai phía	$W_\alpha = \left\{z_0 : \lvert z_0\rvert > z_{\alpha/2}\right\}$	$2\left[1 - \Phi(\lvert z_0\rvert)\right]$
Một phía trên	$W_\alpha = \{z_0 : z_0 > z_\alpha\}$	$1 - \Phi(z_0)$
Một phía dưới	$W_\alpha = \{z_0 : z_0 < -z_\alpha\}$	$\Phi(z_0)$

Bảng 1

Đối thuyết H1	Miền bác bỏ (một mẫu)	Miền bác bỏ (hai mẫu)	Trị số p_v
Hai phía	$W_\alpha = \left\{t_0 : \lvert t_0\rvert > t_{\alpha/2,n-1}\right\}$	$W_\alpha = \left\{t_0 : \lvert t_0\rvert > t_{\alpha/2,df}\right\}$	$2P(T > \lvert t_0\rvert)$
Một phía trên	$W_\alpha = \{t_0 : t_0 > t_{\alpha,n-1}\}$	$W_\alpha = \{t_0 : t_0 > t_{\alpha,df}\}$	$P(T > t_0)$
Một phía dưới	$W_\alpha = \{t_0 : t_0 < -t_{\alpha,n-1}\}$	$W_\alpha = \{t_0 : t_0 < -t_{\alpha,df}\}$	$P(T < t_0)$

Bảng 2

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: $H_0 : \tau_1 = \tau_2 = \cdots = \tau_k = 0$ 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)	$SSB = n\sum_{i=1}^{k}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{k}\frac{y_{i\cdot}^2}{n} - \frac{y_{\cdot\cdot}^2}{N}$	$k-1$	$MSB = \frac{SSB}{k-1}$	
Trong từng nhóm (SSW)	$SSW = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{i\cdot})^2 = SST - SSB$	$k(n-1)$	$MSW = \frac{SSW}{k(n-1)}$	$F = \frac{MSB}{MSW}$
Tổng (SST)	$SST = \sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n}y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{N}$	$kn-1$		

2.5 Hồi quy tuyến tính đơn

Đường hồi quy tuyến tính mẫu Y **theo** X : $y = \hat{\beta}_0 + \hat{\beta}_1 x$. Trong đó: $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{n} x_i y_i - \dfrac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}} = \frac{S_{xy}}{S_{xx}}$, và $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$.

$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - \dfrac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$ và $S_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n} x_i y_i - \dfrac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}$

Hệ số tương quan mẫu : $R_{XY}^2 = \beta_1^2\dfrac{S_{xx}}{SST}$. Trong đó: $SST = \sum_i (y_i - \overline{y})^2$.