TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP.HCM

KHOA CÔNG NGHỆ THÔNG TIN

BTC ÔN THI HỌC KỲ 1 KHÓA 2016

Xác suất Thống kê

(Vũ Lê Thế Anh)

Cập nhật: 02/12/2017

Mục lục

PHÂN PHỐI BIẾN NGẪU NHIÊN

1. Phân phối rời rạc

1.1. Phân phối nhị thức

Ý nghĩa: Biến ngẫu nhiên X có phân phối nhị thức là số phần tử thỏa mãn một điều kiện nào đó trong một tập hợp có n phần tử, với xác suất thỏa mãn điều kiện của mỗi phần tử là độc lập và bằng p.

Ký hiệu: $X \sim B(n, p)$

Hàm xác suất:

$$f(x) = P(X = x) = \begin{cases} C_n^x \, p^x \, (1-p)^{n-x}, & x = 0,1,2,\dots,n \\ 0, & x \text{ khác} \end{cases}$$

Kỳ vọng: $E(X) = np$

Phương sai: $Var(X) = np(1-p)$

Mod: $(n+1)p - 1 \le Mod(X) \le (n+1)p,\ Mod(X) \in \mathbb{N}$

1.2. Phân phối Poisson

Ý nghĩa: Biến ngẫu nhiên X có phân phối Poisson là số lần một sự kiện xảy ra trong một khoảng thời gian (hoặc không gian) cố định. Thường áp dụng cho số người đến một nơi (quán ăn, bưu điện,…), số hạt nho trong bánh mì, số lỗi sai trong văn bản,…

Ký hiệu: $X \sim P(\lambda)$

Hàm xác suất:

$$f(x) = P(X = x) = \begin{cases} e^{-\lambda} \dfrac{\lambda^x}{x!} & x = 0,1,2,\dots \\ 0 & x \text{ khác} \end{cases}$$

Kỳ vọng: $E(X) = \lambda$

Phương sai: $Var(X) = \lambda$

Mod: $\lambda - 1 \le Mod(X) \le \lambda,\ Mod(X) \in \mathbb{N}$

<u>**Xấp xỉ phân phối nhị thức bằng phân phối Poisson:**</u>

Điều kiện: $n \ge 100, p \le 0.1, np \le 20$

Khi đó, $B(n,p) \approx P(\lambda)$ với $\lambda = np$.

2. Phân phối liên tục

2.1. Phân phối đều

Ký hiệu: $X \sim U(a, b)$

Hàm mật độ:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

Hàm phân phối:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & x \in [a, b] \\ 1, & x > b \end{cases}$$

Kỳ vọng: $E(X) = \frac{a+b}{2}$

Phương sai: $Var(X) = \frac{(b-a)^2}{12}$

Mod: $Mod(X) = m, \forall m \in [a, b]$

2.2. Phân phối chuẩn

Ký hiệu: $X \sim N(\mu, \sigma^2)$

Hàm mật độ:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Hàm phân phối:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$$

Để tra bảng: $P(X \leq x) = P\left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

Phép hiệu chỉnh liên tục: $P(X < x) = P(X \leq x) = P(X \leq x + 0.5) = \Phi\left(\frac{x+0.5-\mu}{\sigma}\right)$

Kỳ vọng: $E(X) = \mu$

Phương sai: $Var(X) = \sigma^2$

Mod: $Mod(X) = \mu$

<u>Xấp xỉ phân phối nhị thức bằng phân phối chuẩn:</u>

Điều kiện: $0.1 < p < 0.9, np \geq 5, np(1-p) \geq 5$

Khi đó, $B(n,p) \approx N(\mu, \sigma^2)$ với $\mu = np, \sigma^2 = np(1-p)$.

KHOẢNG TIN CẬY

1. Khoảng tin cậy cho trung bình

<u>Bài toán:</u> Cho tổng thể với trung bình mẫu μ và phương sai σ^2 có thể biết trước hoặc chưa biết. Từ mẫu ngẫu nhiên $(X_1, X_2, \dots, X_n)$ tìm khoảng tin cậy cho μ với độ tin cậy $1-\alpha$ cho trước.

<u>Các bước thực hiện:</u>

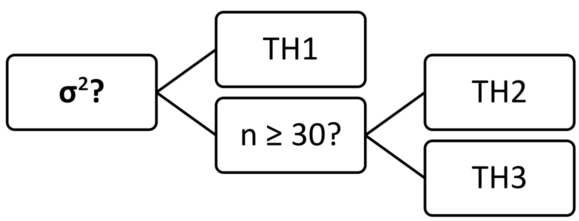
B1: Tìm trung bình mẫu $\bar{x}$ và phương sai mẫu s

B2: Xác định trường hợp áp dụng

Các trường hợp có thể xảy ra:

- TH1: $n \geq 30$ hoặc $n < 30$ nhưng X có phân phối chuẩn, σ^2 đã biết
- TH2: $n \geq 30$, σ^2 chưa biết
- TH3: $n < 30$, X có phân phối chuẩn, σ^2 chưa biết

Cây quyết định:



B3: Tìm phân vị (tra bảng)

<u>TH1, TH2:</u> $z_{1-\alpha/2}$ (phân vị mức $1-\alpha/2$ của luật phân phối chuẩn)

<u>TH3:</u> $t_{1-\alpha/2}^{n-1}$ (phân vị mức $1-\alpha/2$ của luật phân phối Student với bậc tự do $n-1$)

B4: Tính dung sai

Công thức chung

$$\varepsilon = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

<u>TH3:</u> thay $z_{1-\alpha/2}$ bằng $t_{1-\alpha/2}^{n-1}$

σ chưa biết: thay bằng s

Kết luận: Khoảng tin cậy $100(1-\alpha)\%$ cho trung bình của tổng thể là $[\bar{x}-\varepsilon, \bar{x}+\varepsilon]$

2. Khoảng tin cậy cho tỉ lệ

Bài toán: Cho tổng thể X, trong đó tỉ lệ cá thể mang đặc tính $\mathbb{A}$ nào đó là p. Từ mẫu ngẫu nhiên $(X_1, X_2, \dots, X_n)$ tìm khoảng tin cậy cho p với độ tin cậy $1-\alpha$ cho trước.

Các bước thực hiện:

B1: Tìm tỷ lệ mẫu $\hat{p}$

B2: Kiểm tra điều kiện $n\hat{p} \geq 5$ và $n(1-\hat{p}) \geq 5$

B3: Tìm phân vị $z_{1-\alpha/2}$ (phân vị mức $1-\alpha/2$ của luật phân phối chuẩn) bằng bảng tra

B4: Tính dung sai

$$\varepsilon = z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Kết luận: Khoảng tin cậy $100(1-\alpha)\%$ cho tỷ lệ của tổng thể là $[\hat{p}-\varepsilon, \hat{p}+\varepsilon]$

3. Xác định kích cỡ mẫu

Bài toán: Tìm n nhỏ nhất sao cho dung sai (sai số) $\varepsilon \leq \varepsilon_0$ và α cho trước.

Trường hợp tìm KTC cho trung bình: dựa vào công thức dung sai của tùy từng trường hợp mà tìm ra n cụ thể

TH1, TH2: $n \geq \left(z_{1-\alpha/2}\frac{\sigma}{\varepsilon_0}\right)^2$ (dùng s thay thế khi không có σ)

TH3: Tính trước trong trường hợp $n = 29$, tìm ε_{29}.

Nếu $\varepsilon_{29} > \varepsilon_0$, ta cần chọn n lớn hơn => sử dụng $z_{1-\alpha/2}$ thay cho $t_{1-\alpha/2}^{n-1}$ (làm như TH2)

Nếu $\varepsilon_{29} < \varepsilon_0$, thử và chọn n phù hợp (hiếm gặp)

Trường hợp tìm KTC cho tỷ lệ:

Cho trước $\hat{p}$: dựa vào công thức dung sai => $n \geq \left(z_{1-\alpha/2}\frac{\sqrt{\hat{p}(1-\hat{p})}}{\varepsilon_0}\right)^2$

Không cho trước $\hat{p}$: cho $\hat{p} = 0.5$ và làm như trên

KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

1. Kiểm định trong trường hợp một mẫu:

1.1. Kiểm định kỳ vọng:

<u>Bài toán:</u> Cho mẫu ngẫu nhiên $(X_1, X_2, \dots, X_n)$ chọn từ tổng thể X có kỳ vọng μ chưa biết và phương sai σ^2 có thể đã biết. Cho một giả thuyết thống kê gồm giả thuyết không H_0 và đối thuyết H_1. Kiểm định giả thuyết thống kê với mức ý nghĩa α cho trước.

<u>Các trường hợp giả thuyết thống kê:</u>

Giả thuyết không: $H_0: \mu = \mu_0$

Đối thuyết:

(1) $H_1: \mu \neq \mu_0$ (2) $H_1: \mu > \mu_0$ (3) $H_1: \mu < \mu_0$

<u>Các bước kiểm định:</u>

B1: Phát biểu giả thuyết và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính thống kê kiểm định

$$Z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

Trường hợp không cho biết σ: thay thế bằng s

Trường hợp mẫu nhỏ $n < 30$: gọi là T_0 thay vì Z_0 (chỉ là thay đổi ký hiệu, giá trị không bị ảnh hưởng, khác biệt ở bước sau)

B4: Xác định miền bác bỏ

$H_1: \mu \neq \mu_0$	$W_\alpha = \{z_0: \lvert z_0 \rvert > z_{1-\alpha/2}\}$
$H_1: \mu > \mu_0$	$W_\alpha = \{z_0: z_0 > z_{1-\alpha}\}$
$H_1: \mu < \mu_0$	$W_\alpha = \{z_0: z_0 < -z_{1-\alpha}\}$

Trường hợp mẫu nhỏ $n < 30$: thay vì dùng phân phối chuẩn z_{1-x} thì ta dùng phân phối Student bậc tự do $(n-1)$, tức dùng t_{1-x}^{n-1} thay cho z_{1-x} (thay z_0 bằng t_0)

Kết luận:

Giá trị thống kê kiểm định nằm trong miền bác bỏ: bác bỏ H_0, chấp nhận H_1

Giá trị thống kê kiểm định không nằm trong miền bác bỏ: không đủ cơ sở để bác bỏ H_0

1.2. Kiểm định tỷ lệ:

Bài toán: Cho mẫu ngẫu nhiên $(X_1, X_2, \dots, X_n)$ chọn từ tổng thể X có tỉ lệ cá thể có đặc tính $\mathbb{A}$ là p (chưa biết) và một giả thuyết thống kê gồm giả thuyết không H_0 và đối thuyết H_1. Kiểm định giả thuyết thống kê với mức ý nghĩa α cho trước.

Các trường hợp giả thuyết thống kê:

Giả thuyết không: $H_0: p = p_0$

Đối thuyết:

(1) $H_1: p \neq p_0$ (2) $H_1: p > p_0$ (3) $H_1: p < p_0$

Các bước kiểm định:

B1: Phát biểu giả thuyết và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính thống kê kiểm định

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}}$$

B4: Xác định miền bác bỏ

$H_1: p \neq p_0$	$W_\alpha = \{z_0: \lvert z_0 \rvert > z_{1-\alpha/2}\}$
$H_1: p > p_0$	$W_\alpha = \{z_0: z_0 > z_{1-\alpha}\}$
$H_1: p < p_0$	$W_\alpha = \{z_0: z_0 < -z_{1-\alpha}\}$

Kết luận:

Giá trị thống kê kiểm định nằm trong miền bác bỏ: bác bỏ H_0, chấp nhận H_1

Giá trị thống kê kiểm định không nằm trong miền bác bỏ: không đủ cơ sở để bác bỏ H_0

2. Kiểm định trong trường hợp hai mẫu độc lập:

2.1. So sánh hai kỳ vọng:

Bài toán: Cho hai mẫu ngẫu nhiên $(X_1, X_2, \dots, X_n)$ và $(Y_1, Y_2, \dots, Y_m)$ chọn từ tổng thể X có kỳ vọng μ_1 μ_2 chưa biết và phương sai σ_1^2, σ_2^2 có thể đã biết. Cho một giả thuyết thống kê gồm giả thuyết không H_0 và đối thuyết H_1. Kiểm định giả thuyết thống kê với mức ý nghĩa α cho trước.

Các trường hợp giả thuyết thống kê:

Giả thuyết không: $H_0: \mu_1 - \mu_2 = D$

Đối thuyết:

(1) $H_1: \mu_1 - \mu_2 \neq D$		(2) $H_1: \mu_1 - \mu_2 > D$		(3) $H_1: \mu_1 - \mu_2 < D$

Các bước kiểm định:

B1: Phát biểu giả thuyết và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính thống kê kiểm định

Trường hợp 1: mẫu lớn, $n > 30$ và $m > 30$

$$Z_0 = \frac{(\bar{x}_1 - \bar{x}_2) - D}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

Nếu chưa biết σ^2 thì thay bằng phương sai mẫu s^2

Trường hợp 2: mẫu nhỏ, chưa biết phương sai, nhưng có $\sigma_1^2 = \sigma_2^2 = \sigma^2$

Phương sai mẫu gộp:

$$s^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}$$

Thống kê kiểm định:

$$T_0 = \frac{(\bar{x}_1 - \bar{x}_2) - D}{s\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

B4: Xác định miền bác bỏ

$H_1: \mu_1 - \mu_2 \neq D$	$W_\alpha = \left\{z_0: \lvert z_0 \rvert > z_{1-\alpha/2}\right\}$
$H_1: \mu_1 - \mu_2 > D$	$W_\alpha = \{z_0: z_0 > z_{1-\alpha}\}$
$H_1: \mu_1 - \mu_2 < D$	$W_\alpha = \{z_0: z_0 < -z_{1-\alpha}\}$

Trường hợp mẫu nhỏ $n < 30$: thay vì dùng phân phối chuẩn z_{1-x} thì ta dùng phân phối Student bậc tự do $(n+m-2)$, tức dùng t_{1-x}^{n+m-2} thay cho z_{1-x} (thay z_0 bằng t_0)

Kết luận:

Giá trị thống kê kiểm định nằm trong miền bác bỏ: bác bỏ H_0, chấp nhận H_1

Giá trị thống kê kiểm định không nằm trong miền bác bỏ: không đủ cơ sở để bác bỏ H_0

2.2. So sánh hai tỷ lệ

Bài toán: Cho hai mẫu ngẫu nhiên $(X_1, X_2, \dots, X_n)$ và $(Y_1, Y_2, \dots, Y_m)$ chọn từ tổng thể X có tỉ lệ cá thể có đặc tính $\mathbb{A}$ là p_1, p_2 (chưa biết) và một giả thuyết thống kê gồm giả thuyết không H_0 và đối thuyết H_1. Kiểm định giả thuyết thống kê với mức ý nghĩa α cho trước.

Các trường hợp giả thuyết thống kê:

Giả thuyết không: $H_0: p_1 - p_2 = D$

Đối thuyết:

(1) $H_1: p_1 - p_2 \neq D$ (2) $H_1: p_1 - p_2 > D$ (3) $H_1: p_1 - p_2 < D$

Các bước kiểm định:

B1: Phát biểu giả thuyết và đối thuyết

B2: Xác định mức ý nghĩa α

B3: Tính thống kê kiểm định

Trung bình mẫu gộp:

$$\hat{p} = \frac{\hat{p}_1 n + \hat{p}_2 m}{n + m}$$

Thống kê kiểm định

$$Z_0 = \frac{(\hat{p}_1 - \hat{p}_2) - D}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}}$$

B4: Xác định miền bác bỏ

$H_1: p_1 - p_2 \neq D$	$W_\alpha = \{z_0: \lvert z_0 \rvert > z_{1-\alpha/2}\}$
$H_1: p_1 - p_2 \neq D$	$W_\alpha = \{z_0: z_0 > z_{1-\alpha}\}$
$H_1: p_1 - p_2 \neq D$	$W_\alpha = \{z_0: z_0 < -z_{1-\alpha}\}$

Kết luận:

Giá trị thống kê kiểm định nằm trong miền bác bỏ: bác bỏ H_0, chấp nhận H_1

Giá trị thống kê kiểm định không nằm trong miền bác bỏ: không đủ cơ sở để bác bỏ H_0