Chương 4:

LÍ THUYẾT ƯỚC LƯỢNG & KIỂM ĐỊNH GIẢ THIẾT

NHẮC LẠI

- Cho X là biến ngẫu nhiên tác động trên một tổng thể nào đó. Xét mẫu có kích thước n là tập hợp của n biến ngẫu nhiên độc lập X_1 , X_2 ,...,X_n được lập từ BNN X và có cùng luật phân phối xác suất của X.
- Ta tiến hành quan sát, thu thập (cân, đo...) từng biến X_i và nhận được các giá trị cụ thể $X_i = x_i$, khi đó (X_1, X_2, …,X_n) là một mẫu ngẫu nhiên của X.

Ví dụ:

Chiều cao của cây bạch đàn là BNN X có phân phối chuẩn. Đo ngẫu nhiên chiều cao của 5 cây bạch đàn ta được X_1 =3,5m, X_2 =3,2m, X_3 =2,5m, X_4 = 4,1m, X_5 =3m. Khi đó, (X_1, X_2, X_3 , X_4, X_5)= (3,5; 3,2; 2,5;4,1;3) là BNN có luật phân phối chuẩn.

CÁC THAM SỐ ĐẶC TRƯNG CƠ BẢN CỦA BNN

- Trung bình mẫu:

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Phương sai mẫu:

$$S^2 = \frac{(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

- Phương sai mẫu hiệu chỉnh:

$$S_1^2 = \frac{n}{n-1} S^2$$

4.1. BÀI TOÁN ƯỚC LƯỢNG

– ***Ước lượng*** là *phỏng đoán* một giá trị chưa biết bằng cách dựa vào quan sát mẫu.

– Thông thường ta cần ước lượng giá trị trung bình, phương sai, độ lệch chuẩn …

– Có 2 hình thức ước lượng:

a. *Ước lượng điểm:* là đi tìm một giá trị gần đúng nhất với giá trị thực của tham số.

b. *Ước lượng khoảng:* là đi tìm một *khoảng số thực* sao cho xác suất của tham số cần ước lượng rơi vào khoảng đó tương đối lớn.

4.1.1. Ước lượng điểm

– Biến ngẫu nhiên X cần nghiên cứu cho tổng thể, giả sử X có luật phân phối xác suất đã biết (chuẩn, nhị thức,…) nhưng còn phụ thuộc vào tham số θ chưa biết.

– Cần ước lượng tham số θ

– Ước lượng $\hat{\theta}$ của tham số θ gọi là *ước lượng tốt nhất* nếu thoả mãn 3 điều kiện sau:

+ *Không chệch:* $E(\hat{\theta}) = \theta$

+ *Bền vững:*

$$\lim_{n\to+\infty} P\left(\left|\hat{\theta}-\theta\right| < \varepsilon\right) = 1 \quad \text{hay} \quad \hat{\theta} \xrightarrow{n\to+\infty} \theta$$

+ *Có hiệu quả:*

$D(\hat{\theta})$ đạt giá trị nhỏ nhất

Một số kết quả

- $\overline{X}$ là ước lượng tốt nhất của kì vọng E(X).

- Phương sai mẫu hiệu chỉnh S_1^2 là một ước lượng không chệch của phương sai D(X) của X.

4.1.2. Ước lượng khoảng

1. Khoảng tin cậy (khoảng ước lượng), độ tin cậy

– Khoảng (c; d) gọi là ***khoảng tin cậy*** (khoảng ước lượng) của tham số θ, nếu:

$$P(\theta \in (c;d)) = 1 - \alpha = \gamma$$

với α khá bé (thường 0 < α < 0,05)

$1 - \alpha = \gamma$ (gần bằng 1) gọi là ***độ tin cậy***

Chú ý

- Theo định nghĩa độ tin cậy $\gamma = 1 - \alpha$ càng gần bằng 1 và khoảng tin cậy càng ngắn thì ước lượng càng chính xác.

– Trường hợp khoảng tin cậy đối xứng có dạng:

$$(\hat{\theta} - \varepsilon; \hat{\theta} + \varepsilon)$$

- ε gọi là độ chính xác hay sai số của ước lượng.
- 2ε gọi là bề rộng của khoảng.

2. Ước lượng khoảng của kì vọng

Bài toán:

X là BNN có phân phối chuẩn N(a, σ), nhưng kì vọng EX = a chưa biết. Tìm khoảng ước lượng đối xứng của kỳ vọng khi đã biết trung bình mẫu $\overline{X}$ với độ tin cậy γ = 1 - α

Giải:

Khoảng ước lượng đối xứng của kì vọng EX có dạng:

$$(\overline{X}-\varepsilon, \overline{X}+\varepsilon)$$

Có 2 trường hợp sau:

Trường hợp 1: Đã biết σ

Thay $\varepsilon = \frac{\sigma}{\sqrt{n}}\Phi^{-1}(\frac{\gamma}{2})$, suy ra khoảng ước lượng

Trường hợp 2: Chưa biết σ

Khi đó, thay độ lệch chuẩn mẫu hiệu chỉnh S_1 bởi σ

Nếu n > 30, thay $\varepsilon = \frac{S_1}{\sqrt{n}}\Phi^{-1}(\frac{\gamma}{2})$, suy ra khoảng ước lượng.

Nếu n ≤ 30, thay $\varepsilon = \frac{S_1}{\sqrt{n}}T_\alpha(n-1)$, suy ra khoảng ước lượng.

(với $T_\alpha(n-1)$ tra bảng phân phối Student)

Ví dụ

Biết X là độ dài của một loại trục do một máy tiện làm ra là biến ngẫu nhiên có phân phối chuẩn N(a, 2)

Từ một mẫu gồm 64 trục, người ta tính được trung bình mẫu $\overline{X} = 20{,}5cm$

Hãy tìm khoảng tin cậy đối xứng cho độ dài trung bình của trục với độ tin cậy là $\gamma = 0{,}95$, biết $\Phi(1{,}96) = 0{,}475$

Giải

Vì X là biến ngẫu nhiên có phân phối chuẩn N(a, 3) nên σ = 2

Khi đó, khoảng tin cậy đối xứng của kì vọng là:

$$(\overline{X} - \varepsilon, \overline{X} + \varepsilon)$$

Với $\varepsilon = \frac{\sigma}{\sqrt{n}} \Phi^{-1}(\frac{\gamma}{2})$

Ta có $\gamma = 0{,}95 \Rightarrow \Phi^{-1}(\frac{\gamma}{2}) = \Phi^{-1}(0{,}475) = 1{,}96$

$$\Rightarrow \varepsilon = \frac{\sigma}{\sqrt{n}} \Phi^{-1}(\frac{\gamma}{2}) = \frac{2}{8} 1{,}96 = 0{,}49$$

Vậy khoảng tin cậy đối xứng của độ dài trung bình là (20,01; 20,99)

Ví dụ

Đo đường kính của 100 trục máy do 1 nhà máy sản xuất được bảng số liệu sau:

Đường kính (cm)	9,75	9,8	9,85	9,9
Số trục máy	5	37	42	16

Hãy ước lượng khoảng trung bình của trục máy với độ tin cậy 95%, biết đường kính của trục máy là BNN có phân phối chuẩn và $\Phi(1,96) = 0,475$

Chú ý: Xác định độ tin cậy và kích thước mẫu

- Độ tin cậy: $1-\alpha = 2\Phi(\frac{\varepsilon\sqrt{n}}{S_1})$
- Kích thước mẫu: $n = \left[\left(\frac{\Phi^{-1}(\frac{\gamma}{2})S_1}{\varepsilon}\right)^2\right] + 1$

Ví dụ

Đo đường kính của 100 chi tiết do một máy sản xuất ta thu được bảng số liệu:

Đường kính (cm)	9,75	9,80	9,85	9,90
Số chi tiết	5	37	42	16

a.Với độ chính xác 0,006 hãy xác định độ tin cậy?

b. Muốn độ chính xác 0,003; độ tin cậy 95% thì cần kiểm tra bao nhiêu chi tiết.

4.2. BÀI TOÁN KIỂM ĐỊNH GIẢ THIẾT

Ở phần trước ta đã nghiên cứu các tham số đặc trưng của tổng thể trên cơ sở thông tin của mẫu bằng *phương pháp ước lượng*. Phần này tiếp tục nghiên cứu dấu hiệu của tổng thể bằng phương pháp khác là *kiểm định giả thiết thống kê.*

Đứng trước một tổng thể, ta thường có một giả thiết nào đó. Bằng cách sử dụng các kết quả nghiên cứu từ mẫu để khẳng định hay bác bỏ giả thiết ban đầu gọi là kiểm định giả thiết thống kê.

Các giả thiết thống kê thường được kí hiệu là: H

Các loại sai lầm của kiểm định giả thiết:

– Loại 1: Bác bỏ H trong khi H đúng.

– Loại 2: Chấp nhận H trong khi H sai.

1. Phương pháp kiểm định:

Chấp nhận xảy ra sai lầm loại 1 với xác suất không vượt quá α (α gọi là mức ý nghĩa). Với mức ý nghĩa đó, ta sẽ chấp nhận giả thiết H nếu xác suất xảy ra sai lầm loại 2 nhỏ nhất.

2. Các loại kiểm định giả thiết về tham số t của biến ngẫu nhiên:

– Kiểm định 2 phía đối với tham số t:

Giả thiết H: "$t = t_o$" với đối thiết H_1: "$t \neq t_o$"

– Kiểm định phía phải đối với tham số t:

Giả thiết H: "$t = t_o$" với đối thiết H_1: "t > to"

– Kiểm định phía trái đối với tham số t:

Giả thiết H: "$t = t_o$" với đối thiết H_1: "$t < t_o$"

3. Kiểm định giả thiết về giá trị trung bình

Bài toán:

Giả sử tổng thể có trung bình (kỳ vọng) là a (giá trị của a chưa biết). Mẫu có kích thước n, trung bình mẫu $\overline{X}$, độ lệch chuẩn mẫu hiệu chỉnh S_1.

Hãy kiểm định giả thiết H: $a = a_0$ với mức ý nghĩa α.

(a_0 là giá trị cụ thể nào đó đã biết).

Kiểm định giả thiết H: a = a_0 với đối thiết H_1 : a $\neq$ a_0

Bước 1: Tính $Z_\alpha = \Phi^{-1}\left(\frac{1-\alpha}{2}\right)$

Bước 2: Tính

$$Z_0 = \frac{|\overline{X} - a_0|}{S_1} \cdot \sqrt{n}$$

Bước 3: Kết luận

-Nếu $Z_0 \leq Z_\alpha$ thì chấp nhận H

- Nếu $Z_0 > Z_\alpha$ thì bác bỏ H và chấp nhận H_1 với mức ý nghĩa α.

Ví dụ:

Điểm trung bình môn toán năm học trước là 5,72. Năm học này, theo dõi 100 sinh viên thu được bảng số liệu sau:

Điểm	3	4	5	6	7	8	9
Số SV	3	5	27	43	12	6	4

Hãy kiểm định giả thiết: Phải chăng điểm trung bình môn toán năm nay đã có sự thay đổi so với năm trước với mức ý nghĩa 1%?