Chương 4

ƯỚC LƯỢNG THAM SỐ CỦA ĐẠI LƯỢNG NGẪU NHIÊN

Giả sử đại lượng ngẫu nhiên X có tham số θ chưa biết. Ước lượng tham số θ là dựa vào mẫu ngẫu nhiên $W_x = (X_1, X_2, \ldots, X_n)$ ta đưa ra thống kê $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ để ước lượng (dự đoán) θ.

Có 2 phương pháp ước lượng:

i) Ước lượng điểm: chỉ ra $\theta = \theta_0$ nào đó để ước lượng θ.

ii) Ước lượng khoảng: chỉ ra một khoảng (θ_1, θ_2) chứa θ sao cho $P(\theta_1 < \theta < \theta_2) = 1 - \alpha$ cho trước ($1 - \alpha$ gọi là độ tin cậy của ước lượng).

1. CÁC PHƯƠNG PHÁP ƯỚC LƯỢNG ĐIỂM

1.1 Phương pháp hàm ước lượng

• Mô tả phương pháp

Giả sử cần ước lượng tham số θ của đại lượng ngẫu nhiên X. Từ X ta lập mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$.

Chọn thống kê $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$. Ta gọi $\hat{\theta}$ là *hàm ước lượng* của X.

Thực hiện phép thử ta được mẫu cụ thể $w_x = (x_1, x_2, \ldots, x_n)$. Khi đó ước lượng điểm của θ là giá trị $\theta_0 = \hat{\theta}(x_1, x_2, \ldots, x_n)$.

a) Ước lượng không chệch

□ **Định nghĩa 1** *Thống kê $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ được gọi là ước lượng không chệch của tham số θ nếu $E(\hat{\theta}) = \theta$.*

⊙ **Ý nghĩa**

Giả sử $\hat{\theta}$ là ước lượng không chệch của tham số θ. Ta có

$$E(\hat{\theta} - \theta) = E(\hat{\theta}) - E(\theta) = \theta - \theta = 0$$

Vậu ước lượng không chệch là ước lượng có sai số trung bình bằng 0.

⊕ **Nhận xét**

i) Trung bình của mẫu ngẫu nhiên $\overline{X}$ là ước lượng không chệch của trung bình của tổng thể $\theta = E(X) = m$ vì $E(\overline{X}) = m$.

ii) Phương sai điều chỉnh của mẫu ngẫu nhiên S'^2 là ước lượng không chệch của phương sai của tổng thể σ^2 vì $E(S'^2) = \sigma^2$.

• **Ví dụ 1** *Chiều cao của 50 cây lim được cho bởi*

Khoảng chiều cao (mét)	*số cây lim*	x_i^0	u_i	n_iu_i	$n_iu_i^2$
$[6,25 - 6,75)$	1	6,5	-4	-4	16
$[6,75 - 7,25)$	2	7,0	-3	-6	18
$[7,25 - 7,75)$	5	7,5	-2	-10	20
$[7,75 - 8,25)$	11	8	-1	-11	11
$[8,25 - 8,75)$	18	8,5	0	0	0
$[8,75 - 9,25)$	9	9	1	9	9
$[9,25 - 9,75)$	3	9,5	2	6	12
$[9,75 - 10,2)$	1	10	3	3	9
$\sum$	50			-13	95

Gọi X là chiều cao của cây lim

a) Hãy chỉ ra ước lượng điểm cho chiều cao trung bình của các cây lim.
b) Hãy chỉ ra ước lượng điểm cho độ tản mát của các chiều cao cây lim so với chiều cao trung bình.
c) Gọi $p = P(7,75 \leq X \leq 8,75)$. Hãy chỉ ra ước lượng điểm cho p.

Giải

Ta lập bảng tính cho $\overline{x}$ và s^2.

Thực hiện phép đổi biến $u_i = \dfrac{x_i^0 - 8,5}{0,5}$ $\quad (x_0 = 8,5;\ h = 0,5)$

Ta có $\overline{u} = -\frac{13}{50} = -0,26$. Suy ra

$\overline{x} = 8,5 + 0,5.(-0,26) = 8,37$

$s^2 = (0,5)^2.\left[\dfrac{95}{50} - (-0,26)^2\right] = 0,4581 \sim (0,68)^2.$

a) Chiều cao trung bình được ước lượng là 8,37 mét.

b) Độ tản mát được ước lượng là $s = 0,68$ mét hoặc $\hat{s} = \sqrt{\frac{50}{50-1}0,4581} \sim 0,684$

c) Trong 50 quan sát đã cho có $11 + 18 = 29$ quan sát cho chiều cao lim thuộc khoảng $[7,5 - 8,5)$

Vậy ước lượng điểm cho p là $p^* = \frac{29}{50} = 0,58$.

b) Ước lượng hiệu quả

⊕ **Nhận xét** Giả sử $\hat{\theta}$ là ước lượng không chệch của tham số θ. Theo bất đẳng thức Tchebychev ta có

$$P(|\hat{\theta} - E(\hat{\theta})| < \varepsilon) > 1 - \frac{Var(\hat{\theta})}{\varepsilon^2}$$

Vì $E(\hat{\theta}) = \theta$ nên $P(|\hat{\theta} - \theta| < \varepsilon) > 1 - \dfrac{Var(\hat{\theta})}{\varepsilon^2}$.

Ta thấy nếu $Var(\hat{\theta})$ càng nhỏ thì $P(|\hat{\theta} - \theta| < \varepsilon)$ càng gần 1. Do đó ta sẽ chọn $\hat{\theta}$ với $Var(\hat{\theta})$ nhỏ nhất.

□ **Định nghĩa 2** *Ước lượng không chệch $\hat{\theta}$ được gọi là ước lượng có hiệu quả của tham số θ nếu $Var(\hat{\theta})$ nhỏ nhất trong các ước lượng của θ.*

⊙ **Chú ý** Người ta chứng minh được rằng nếu $\hat{\theta}$ là ước lượng hiệu quả của θ thì phương sai của nó là

$$Var(\hat{\theta}) = \frac{1}{n.E(\frac{\partial lnf(x,\theta)}{\partial \theta})^2} \tag{4.1}$$

trong đó $f(x, \theta)$ là hàm mật độ xác suất của đại lượng ngẫu nhiên gốc. Mọi ước lượng không chệch θ luôn có phương sai lớn hơn $Var(\hat{\theta})$ trong (4.1). Ta gọi (4.1) là *giới hạn Crame-Rao*.

⊕ **Nhận xét** Nếu đại lượng ngẫu nhiên gốc $X \in N(\mu, \frac{\sigma^2}{n})$ thì trung bình mẫu $\overline{X}$ là ước lượng hiệu quả của kỳ vọng $E(X) = \mu$.

Thật vậy, ta biết $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i \in N(\mu, \dfrac{\sigma^2}{n})$

Mặt khác do X có phân phối chuẩn nên nếu $f(x, \mu)$ là hàm mật độ của X_i thì

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

Ta có $\dfrac{\partial}{\partial \mu} lnf(x, \mu) = \dfrac{x - \mu}{\sigma^2}$.

Suy ra $nE\left[\dfrac{\partial lnf(x, \mu)}{\partial \mu}\right]^2 = nE\left(\dfrac{x-\mu}{\sigma^2}\right)^2 = \dfrac{n}{\sigma^2}$. Do đó $Var(\overline{X})$ chính bằng nghịch đảo σ^2/n.

Vậy $\overline{X}$ là ước lượng hiệu quả của μ.

c) Ước lượng vững

□ **Định nghĩa 3** *Thống kê $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ được gọi là ước lượng vững của tham số θ nếu $\forall \varepsilon > 0$ ta có*

$$\lim_{n \to \infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$$

⊙ Điều kiện đủ của ước lượng vững

Nếu $\hat{\theta}$ là ước lượng không chệch của θ và $\lim_{n\to\infty} Var(\hat{\theta}) = 0$ thì $\hat{\theta}$ là ước lượng vững của θ.

1.2 Phương pháp ước lượng hợp lý tối đa

Giả sử $W_X = (X_1, X_2, \ldots, X_n)$ là mẫu ngẫu nhiên được tạo nên từ đại lượng ngẫu nhiên X có mẫu cụ thể $w_x = (x_1, x_2, \ldots, x_n)$ và $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$.

Xét hàm hàm hợp lý $L(x_1, \ldots, x_n, \theta)$ của đối số θ xác định như sau:

- Nếu X rời rạc:

$$L(x_1, \ldots, x_n, \theta) = P(X_1 = x_1/\theta, \ldots, X_n = x_n/\theta) \quad (4.2)$$

$$= \prod_{i=1}^{n} P(X_i = x_i/\theta) \quad (4.3)$$

$L(x_1, \ldots, x_n, \theta)$ là xác suất để ta nhận được mẫu cụ thể $W_x = (x_1, \ldots, x_n)$

- Nếu X liên tục có hàm mật độ xác suất $f(x, \theta)$

$$L(x_1, \ldots, x_n, \theta) = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta)$$

$L(x_1, x_2, \ldots, x_n, \theta)$ là mật độ của xác suất tại điểm $w_x(x_1, x_2, \ldots, x_n)$

Giá trị $\theta_0 = \hat{\theta}(x_1, x_2, \ldots, x_n)$ được gọi là ước lượng hợp lý tối đa nếu ứng với giá trị này của θ hàm hợp lý đạt cực đại.

⊙ Phương pháp tìm

Vì hàm L và lnL đạt cực đại tại cùng một giá trị θ nên ta xét lnL thay vì xét L.

Bước 1: Tìm $\dfrac{\partial lnL}{\partial \theta}$

Bước 2: Giải phương trình $\dfrac{\partial lnL}{\partial \theta}$ (Phương trình hợp lý)

Giả sử phương trình có nghiệm là $\theta_0 = \hat{\theta}(x_1, x_2, \ldots, x_n)$

Bước 3: Tìm đạo hàm cấp hai $\dfrac{\partial^2 lnL}{\partial \theta}$

Nếu tại θ_0 mà $\dfrac{\partial^2 lnL}{\partial \theta} < 0$ thì lnL đạt cực đại. Khi đó $\theta_0 = \hat{\theta}(x_1, x_2, \ldots, x_n)$ là ước lượng điểm hợp lý tối đa của θ.

2. PHƯƠNG PHÁP KHOẢNG TIN CẬY

2.1 Mô tả phương pháp

Giả sử tổng thể có tham số θ chưa biết. Ta tìm khoảng (θ_1, θ_2) chứa θ sao cho $P(\theta_1 < \theta < \theta_2) = 1 - \alpha$ cho trước.

Từ đại lượng ngẫu nhiên gốc X lập mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$. Chọn thống kê $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ có phân phối xác suất xác định dù chưa biết θ.

Với α_1 khá bé ($\alpha_1 < \alpha$) ta tìm được phân vị θ_{α_1} của $\hat{\theta}$ (tức là $P(\hat{\theta} < \theta_{\alpha_1}) = \alpha_1$).

Với α_2 mà $\alpha_1 + \alpha_2 = \alpha$ khá bé (thường lấy $\alpha \leq 0,05$) ta tìm được phân vị $\theta_{1-\alpha_2}$ của $\hat{\theta}$ (tức là $P(\hat{\theta} < \theta_{1-\alpha_2}) = 1 - \alpha_2$).

Khi đó

$$P(\theta_{\alpha_1} \leq \hat{\theta} \leq \theta_{1-\alpha_2}) = P(\hat{\theta} < \theta_{1-\alpha_2}) - P(\hat{\theta} < \theta_{\alpha_1}) = 1 - \alpha_2 - \alpha_1 = 1 - \alpha \qquad (*)$$

Từ (*) ta giải ra được θ. Khi đó (*) được đưa về dạng $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha$.

Vì xác suất $1 - \alpha$ gần bằng 1, nên biến cố $(\hat{\theta}_1 < \theta < \hat{\theta}_2)$ hầu như xảy ra. Thực hiện một phép thử đối với mẫu ngẫu nhiên W_X ta thu được mẫu cụ thể $w_x = (x_1, x_2, \ldots, x_n)$. Từ mẫu cụ thể này ta tính được giá trị $\theta_1 = \hat{\theta}_1(x_1, x_2, \ldots, x_n)$, $\theta_2 = \hat{\theta}_2(x_1, x_2, \ldots, x_n)$.

Vậy với $1 - \alpha$ cho trước, qua mẫu cụ thể w_x ta tìm được khoảng (θ_1, θ_2) chứa θ sao cho $P(\theta_1 < \theta < \theta_2) = 1 - \alpha$.

- Khoảng (θ_1, θ_2) được gọi là khoảng tin cậy.
- $1 - \alpha$ được gọi là độ tin cậy của ước lượng.
- $|\theta_2 - \theta_1|$ được gọi là độ dài khoảng tin cậy.

2.2 Ước lượng trung bình

Giả sử trung bình của tổng thể $E(X) = m$ chưa biết. Ta tìm khoảng (m_1, m_2) chứa m sao cho $P(m_1 < m < m_2) = 1 - \alpha$, với $1 - \alpha$ là độ tin cậy cho trước.

i) Trường hợp 1

$$\begin{cases} \text{Biết } Var(X) = \sigma^2 \\ n \geq 30 \quad \text{hoặc} \quad (n < 30 \text{ nhưng X có phân phối chuẩn}) \end{cases}$$

Chọn thống kê

$$U = \frac{(\overline{X} - m)\sqrt{n}}{\sigma} \qquad (4.4)$$

Ta thấy $U \in N(0, 1)$.

Chọn cặp α_1 và α_2 sao cho $\alpha_1 + \alpha_2 = \alpha$ và tìm các phân vị

$$P(U < u_{\alpha_1}) = \alpha_1, \quad P(U < u_{\alpha_2}) = 1 - \alpha_2$$

Do phân vị chuẩn có tính chất $u_{\alpha_1} = -u_{1-\alpha_1}$ nên

$$P(-u_{1-\alpha_1} < U < u_{1-\alpha_2}) = 1 - \alpha \tag{4.5}$$

Dựa vào (4.4) và giải hệ bất phương trình trong (4.5) ta được

$$\overline{X} - \frac{\sigma}{\sqrt{n}} u_{1-\alpha_2} < m < \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha_1}$$

Để được khoảng tin cậy đối xứng ta chọn $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ và đặt $\gamma = 1 - \frac{\alpha}{2}$ thì

$$\overline{X} - \frac{\sigma}{\sqrt{n}} u_\gamma < m < \overline{X} + \frac{\sigma}{\sqrt{n}} u_\gamma$$

Tóm lại, ta tìm được khoảng tin cậy $(\overline{x} - \varepsilon, \overline{x} + \varepsilon)$, trong đó

* $\overline{x}$ là trung bình của mẫu ngẫu nhiên.

* $\boxed{\varepsilon = u_\gamma \frac{\sigma}{\sqrt{n}}}$ (độ chính xác) với u_γ là phân vị chuẩn mức $\gamma = 1 - \frac{\alpha}{2}$

• **Ví dụ 2** *Khối lượng sản phẩm là đại lượng ngẫu nhiên X có phân phối chuẩn với độ lệch tiêu chuẩn $\sigma = 1$. Cân thử 25 sản phẩm ta thu được kết quả sau*

X (khối lượng)	*18*	*19*	*20*	*21*
n_i (số lượng	*3*	*5*	*15*	*2*

Hãy ước lượng trung bình khối lượng của sản phẩm với độ tin cậy 95 %.

Giải

x_i	n_i	$x_i n_i$
18	3	54
19	5	95
20	15	300
21	2	42
$\sum$	25	491

Ta có $\overline{x} = \frac{491}{25} = 19,64kg$.

Độ tin cậy $1 - \alpha = 0,95 \implies \alpha = 0,025 \implies \gamma = 1 - \frac{\alpha}{2} = 0,975$ Ta tìm được phân vị chuẩn $u_\gamma = u_{0,975} = 1,96$. Do đó

$$\varepsilon = u_{0,975} \frac{1}{\sqrt{25}} = 1,96.\frac{1}{5} = 0.39$$

$$x_1 = \overline{x} - \varepsilon = 19,6 - 0,39 = 19,25$$

$$x_2 = \overline{x} + \varepsilon = 19,6 + 0,39 = 20,03$$

Vậy khoảng tin cậy là $(19,25; 20,03)$.

ii) Trường hợp 2

$$\begin{cases} \sigma^2 \text{ chưa biết} \\ n \geq 30 \end{cases}$$

Trường hợp này kích thước mẫu lớn ($n \geq 30$) có thể dùng ước lượng của S'^2 thay cho σ^2 chưa biết ($E(S'^2) = \sigma^2$), ta tìm được khoảng tin cậy $(\overline{x} - \varepsilon, \overline{x} + \varepsilon)$ trong đó

* $\overline{x}$ là trung bình của mẫu cụ thể.

* $\boxed{\varepsilon = u_\gamma \frac{s'}{\sqrt{n}}}$ với u_γ là phân vị chuẩn mức $\gamma = 1 - \frac{\alpha}{2}$ và s' là độ lệch tiêu chuẩn điều chỉnh của mẫu cụ thể.

• **Ví dụ 3** *Người ta tiến hành nghiên cứu ở một trường đại học xem trong một tháng trung bình một sinh viên tiêu hết bao nhiêu tiền gọi điện thoại. Lấy một mẫu ngẫu nhiên gồm 59 sinh viên thu được kết quả sau:*

14	18	22	30	36	28	42	79	36	52	15	47
95	16	27	111	37	63	127	23	31	70	27	11
30	147	72	37	25	7	33	29	35	41	48	15
29	73	26	15	26	31	57	40	18	85	28	32
22	36	60	41	35	26	20	58	33	23	35	

Hãy ước lượng khoảng tin cậy 95% cho số tiền gọi điện thoại trung bình hàng tháng của một sinh viên.

Giải

Từ các số liệu đã cho, ta có

$$n = 59; \quad \overline{x} = 41,05; \quad s' = 27,99$$

Độ tin cậy $1 - \alpha = 0,95 \implies 1 - \frac{\alpha}{2} = 0,975$. Tra bảng phân vị chuẩn ta có $u_{0,975} = 1,96$.

Do đó $\varepsilon = 1,96.\frac{27,99}{\sqrt{59}} = 7,13$.

$$\overline{x} - 7,13 = 33,92; \quad \overline{x} + 7,13 = 48,18$$

Vậy khoảng tin cậy của ước lượng là (33,92; 48,18).

iii) Trường hợp 3

$$\begin{cases} \sigma^2 \text{ chưa biết} \\ n < 30 \text{ và X có phân phối chuẩn} \end{cases}$$

Chọn thống kê $T = \dfrac{(\overline{X} - m)\sqrt{n}}{S'} \quad \in \quad T(n-1)$.

Ta tìm được khoảng tin cậy $(\overline{x}-\varepsilon, \overline{x}+\varepsilon)$ trong đó $\boxed{\varepsilon = t_\gamma \dfrac{S'}{\sqrt{n}}}$

với t_γ là phân vị Student mức $\gamma = 1-\frac{\alpha}{2}$ với $n-1$ bậc tụ do và s' là độ lệch tiêu chuẩn điều chỉnh của mẫu cụ thể.

• **Ví dụ 4** *Dioxide Sulfur và Oxide Nitrogen là các hóa chất được khai thác từ lòng đất. Các chất này được gió mang đi rất xa, kết hợp thành acid và rơi trở lại mặt đất tạo thành mưa acid. Người ta đo độ đậm đặc của Dioxide Sulfur ($\mu g/m^3$) trong khu rừng Bavarian của nước Đức. Số liệu cho bởi bảng dưới đây:*

52,7	*43,9*	*41,7*	*71,5*	*47,6*	*55,1*
62,2	*56,5*	*33,4*	*61,8*	*54,3*	*50,0*
45,3	*63,4*	*53,9*	*65,5*	*66,6*	*70,0*
52,4	*38,6*	*46,1*	*44,4*	*60,7*	*56,4*

Hãy ước lượng độ đậm đặc trung bình của Dioxide Sulsfur với độ tin cậy 95%.

Giải

Ta tính được $\overline{x} = 53,92\mu g/m^3$, $s' = 10,07\mu g/m^3$.

Độ tin cậy $1-\alpha = 0,95 \implies \alpha = 0,025 \implies 1-\frac{\alpha}{2} = 0,975$. Tra bảng phân vị student mức 0,975 bậc $n-1 = 23$ ta được $t_{23;0,975} = 2,069$.

Do đó $\varepsilon = 2,069\frac{10,07}{\sqrt{24}} = 4,25$.

$$\overline{x}-\varepsilon = 53,92-4,25 = 49,67, \quad \overline{x}+\varepsilon = 53,92+4,25 = 58,17$$

Vậy khoảng tin cậy là (49,67; 58,17).

Người ta biết được nếu độ đậm đặc của Dioxide Sulfur trong một khu vực lớn hơn $20\mu g/m^3$ thì môi trường trong khu vực bị phá hoại bởi mưa acid. Qua ví dụ này các nhà khoa học đã tìm ra được nguyên nhân rừng Bavarian bị phá hoại trầm trọng năm 1983 là do mưa acid .

⊙ **Chú ý** (Xác định kích thước mẫu)

Nếu muốn độ tin cậy $1-\alpha$ và độ chính xác ε đạt ở mức cho trước thì ta cần xác định kích thước n của mẫu.

i) *Trường hợp biết* $Var(X) = \sigma^2$:

Từ công thức $\varepsilon = u_\gamma^2 \frac{\sigma}{\sqrt{n}}$ ta suy ra

$$\boxed{n = u_\gamma^2 \frac{\sigma^2}{\varepsilon^2}}$$

ii) *Trường hợp chưa biết* σ^2:

Dựa và mẫu cụ thể đã cho (nếu chưa có mẫu thì ta có thể tiến hành lấy mẫu lần đầu với kích thước $n_1 \geq 30$) để tính s'^2. Từ đó xác định được

$$\boxed{n = u_\gamma^2 \frac{s'^2}{\varepsilon^2}}$$

Kích thước mẫu n phải là số nguyên. Nếu khi tính n theo các công thức trên được giá trị không nguyên thì ta lấy phần nguyên của nó cộng thêm với 1.

Tức là $n = \left[u_\gamma^2 \frac{\sigma^2}{\varepsilon^2}\right] + 1$ hoặc $n = \left[u_\gamma^2 \frac{s'^2}{\varepsilon^2}\right] + 1$.

2.3 Ước lượng tỷ lệ

Giả sử tổng thể được chia ra làm hai loại phần tử. Tỷ lệ phần tử có tính chất A là p chưa biết. Ước lượng tỷ lệ là chỉ ra khoảng (f_1, f_2) chứa p sao cho $P(f_1 < p < f_2) = 1-\alpha$.

Để cho việc giải bài toán được đơn giản, ta chọn mẫu với kích thước n khá lớn.

Gọi X là số phần tử có tính chất A khi lấy ngẫu nhiên một phần tử từ tổng thể thì X là đại lượng ngẫu nhiên có phân phối xác suất

X	0	1
P	1-p	p

Gọi X_i $(i = \overline{1,n})$ là số phần tử có tính chất A trong lần lấy thứ i.

Ta có $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ chính là tần suất ước lượng điểm của $p = E(X)$. Mặt khác, theo chương 2, $n\overline{X}$ có phân phối nhị thức $B(n,p)$. Từ đó $E(\overline{X}) = p$ và $Var(\overline{X}) = \frac{p(1-p)}{n}$.

Chọn thống kê $U = \frac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}}$, trong đó f là tỷ lệ các phần tử của mẫu có tính chất A.

Khi n khá lớn thì $U \in N(0,1)$. Giải quyết bài toán tương tự như ở ước lượng trung bình, thay $\overline{X}$ bởi f, σ^2 bởi $f(1-f)$... ta được

$$f - u_\gamma\sqrt{\frac{f(1-f)}{n}} < p < f + u_\gamma\sqrt{\frac{f(1-f)}{n}}$$

Tóm lại, ta xác định được khoảng tin cậy $(f_1, f_2) = (f-\varepsilon, f+\varepsilon)$, trong đó

f là tỷ lệ các phần tử của mẫu có tính chất A

$$\boxed{\varepsilon = u_\gamma\sqrt{\frac{f(1-f)}{n}}} \quad \text{(độ chính xác)} \tag{4.6}$$

với u_γ là phân vị chuẩn mức $1-\frac{\alpha}{2}$.

Từ (4.6) ta có

$$\boxed{u_\gamma = \frac{\varepsilon\sqrt{n}}{\sqrt{f(1-f)}}}$$

$$\boxed{n = u^2_{1-\frac{\alpha}{2}}\frac{f(1-f)}{\varepsilon^2}}$$

⊙ **Chú ý** Ta có thể tìm khoảng tin cậy của p bằng cách khác như sau:

Từ khoảng tin cậy của p:

$$\left(f - u_\gamma\sqrt{\frac{p(1-p)}{n}} < p < f + u_\gamma\sqrt{\frac{p(1-p)}{n}}\right) \quad \text{hay} \quad \left(|f-p| < u_\gamma\sqrt{\frac{p(1-p)}{n}}\right)$$

Giải bất phương trìnhnày ta tìm được

$$p_1 = \frac{nf + 0,5u_\gamma^2 - \sqrt{0,25u_\gamma^2 - nf(1-f)}}{n+u_\gamma^2}, \quad p_2 = \frac{nf + 0,5u_\gamma^2 + \sqrt{0,25u_\gamma^2 - nf(1-f)}}{n+u_\gamma^2}$$

Khi đó (p_1, p_2) là khoảng tin cậy của p với độ tin cậy $1-\alpha$.

• **Ví dụ 5** *Kiểm tra 100 sản phẩm trong lô hàng thấy có 20 phế phẩm.*

i) Hãy ước lượng tỷ lệ phế phẩm có độ tin cậy 99 %.

ii) Nếu độ chính xác $\varepsilon = 0,04$ thì độ tin cậy của ước lượng là bao nhiêu?

iii) Nếu muốn có độ tin cậy 99% và độ chính xác 0,04 thì phải kiểm tra bao nhiêu sản phẩm?

Giải

i) $n = 100, \quad f = \frac{20}{100} = 0.2$

Xét $\quad U = \frac{(f-p)\sqrt{100}}{\sqrt{pq}} \quad \in N(0,1).$

Ta có

$$1-\alpha = 0,99 \quad \Longrightarrow \alpha = 0,01 \quad \Longrightarrow 1-\frac{\alpha}{2} = 1 - 0,005 = 0,995$$

$$\varepsilon = u_{0,995}\frac{\sqrt{0,2.0,8}}{\sqrt{100}} = 2,58.\frac{0,4}{10} = 0,1$$

$$f_1 = f - \varepsilon = 0,2 - 0,1 = 0,1$$

$$f_2 = f + \varepsilon = 0,2 + 0,1 = 0,3$$

Vậy khoảng tin cậy là $(0,1;0,3)$.

ii) $u_{1-\frac{\alpha}{2}} = \dfrac{0,04.\sqrt{100}}{\sqrt{0,2.0,8}} = 1$

Tìm được

$$1-\frac{\alpha}{2} = 0,84 \quad \Longrightarrow \quad 1-\alpha = 0,68$$

Vậy độ tin cậy là 68%.

iii)$1-\alpha = 0,99 \quad \Longrightarrow \quad \alpha = 0,01 \quad \Longrightarrow \quad 1-\frac{\alpha}{2} = 0,995$. Tìm được $u_{0,995} = 2,576$.

Do đó

$$n \approx \frac{(2,576)^2.0,2.0,8}{(0,04)^2} = 6,635.100 = 663,5$$

Vậy $n = 664$

2.4 Ước lượng phương sai

Giả sử đại lượng ngẫu nhiên X có phân phối chuẩn với phương sai $Var(X) = \sigma^2$ chưa biết. Cho $0 < \alpha < 0.05$. Ước lượng phương sai $Var(X)$ là chỉ ra khoảng (σ_1^2, σ_2^2) chứa σ^2 sao cho $P(\sigma_1^2 < \sigma^2 < \sigma_2^2) = 1-\alpha$.

Từ X lập mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ và xét các trường hợp

a) Biết $E(X) = \mu$.

Chọn thống kê $\chi^2 = \displaystyle\sum_{i=1}^{n} \frac{(X_i-\mu)^2}{\sigma^2}$

Ta thấy χ^2 có phân phối "khi-bình phương" với n bậc tự do.

Chọn α_1 và α_2 khá bé sao cho $\alpha_1 + \alpha_2 = \alpha$. Ta tìm được các phân vị $\chi^2_{\alpha_1}$ và $\chi^2_{1-\alpha_2}$ thỏa mãn

$$P(\chi^2_{\alpha_1} < \chi^2 < \chi^2_{1-\alpha_2}) = 1-\alpha \tag{4.7}$$

Thay biểu thức của χ^2 vào (4.7) và giải ra ta được

$$\frac{\sum(X_i-\mu)^2}{\chi^2_{1-\alpha_2}} < \sigma^2 < \frac{\sum(X_i-\mu)^2}{\chi^2_{\alpha_1}}$$

Chọn $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ thì

$$\frac{\sum(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}} < \sigma^2 < \frac{\sum(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}} \tag{4.8}$$

Với mẫu cụ thể $w_x = (x_1, x_2, \ldots, x_n)$, tính các tổng $\sum(x_i-\mu)^2$ và dựa vào (4.8) ta tìm được khoảng tin cậy (σ_1^2, σ_2^2), trong đó

$$\boxed{\begin{array}{l} \sigma_1^2 = \dfrac{\sum(x_i-\mu)^2 n_i}{\chi^2_{n,1-\frac{\alpha}{2}}} \\ \sigma_2^2 = \dfrac{\sum(x_i-\mu)^2 n_i}{\chi^2_{n,\frac{\alpha}{2}}} \end{array}}$$

với

$\chi^2_{n,1-\frac{\alpha}{2}}$ là phân vị "khi−bình phương" mức $1-\frac{\alpha}{2}$ với n bậc tự do.

$\chi^2_{n,\frac{\alpha}{2}}$ là phân vị "khi−bình phương" mức $\frac{\alpha}{2}$ với n bậc tự do.

b) Chưa biết $E(X)$.

Chọn thống kê $\chi^2 = \dfrac{(n-1)S^2}{\sigma^2}$

Thống kê này có phân phối "khi−bình phương với $n-1$ bậc tự do. Tương tự như trên ta tìm được khoảng tin cậy (σ_1^2, σ_2^2) với

$$\boxed{\sigma_1^2 = \frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}};\quad \sigma_2^2 = \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}}$$

• **Ví dụ 6** *Mức hao phí nhiên liệu cho một đơn vị sản phẩm là đại lượng ngẫu nhiên có phân phối chuẩn. Xét trên 25 sản phẩm ta thu được kết quả sau:*

X	*19,5*	*20*	*20,5*
n_i	*5*	*18*	*2*

Hãy ước lượng phương sai với độ tin cậy 90 % trong các trường hợp sau:

i) Biết kỳ vọng $\mu = 20g$.

ii) Chưa biết kỳ vọng.

Giải

i) Biết $\mu = 20g$.

x_i	n_i	x_i-20	$(x_i-20)^2$	$(x_i-20)^2n_i$
19,5	5	-0,5	0,25	1,25
20	18	0	0	0
20,5	2	0,5	0,25	0,5
$\sum$	n=25			1,75

Độ tin cậy $1-\alpha = 0,9 \implies \alpha = 0,1 \implies \dfrac{\alpha}{2} = 0,05 \implies 1-\dfrac{\alpha}{2} = 0.95$

Tra bảng phân vị χ^2 với $n = 25$ bậc tự do ta được

$$\chi^2_{25;0,05} = 14,6; \qquad \chi^2_{25;0,95} = 37,7$$

Do đó

$$\sigma_1^2 = \frac{\sum(x_i - 20)^2 n_i}{\chi^2_{25;0,95}} = \frac{1,75}{37,7} = 0,046$$

$$\sigma_2^2 = \frac{\sum(x_i - 20)^2 n_i}{\chi^2_{25;0,05}} = \frac{1,75}{14,6} = 0,12$$

Vậy khoảng tin cậy là $(0,046;\ 0,12)$.

ii) Khi chưa biết kỳ vọng ta tìm $s'^2 = 0,0692$.

Tra bảng phân vị khi bình phương với bậc tự do $n - 1 = 24$.

$$\chi^2_{0,05} = 13,85; \quad \chi^2_{0,95} = 36,4$$

và tính

$$\sigma_1^2 = \frac{24 s'^2}{\chi^2_{0,95}} = \frac{24 \times 0,0692}{36,4} = 0,046$$

$$\sigma_2^2 = \frac{24 s'^2}{\chi^2_{0,05}} = \frac{24 \times 0,0692}{13,85} = 0,12$$

Vậy khoảng tin cậy là $(0,046;\ 0,12)$.

3. BÀI TẬP

1. Một mẫu các trọng lượng tương ứng là 8,3; 10,6; 9,7; 8,8; 10,2 và 9,4 kg. Xác định ước lượng không chệch của

   a) trung bình của tổng thể,
   b) phương sai của tổng thể.

2. Một mẫu độ đo 5 đường kính của quả cầu là 6,33; 6,37; 6,36; 6,32 và 6,37cm. Xác định ước lượng không chệch của trung bình và phương sai của đường kính quả cầu.

3. Để xác định độ chính xác của một chiếc cân tạ không có sai số hệ thống, người ta tiến hành 5 lần cân độc lập (cùng một vật), kết quả như sau:

   $$94,1 \quad 94,8 \quad 96,0 \quad 95,2\ kg$$

   Xác định ước lượng không chệch của phương sai số đo trong hai trường hợp:

   a) biết khối lượng vật cân là 95kg;
   b) không biết khối lượng vật cân.

4. Đường kính của một mẫu ngẫu nhiên của 200 viên bi được sản xuất bởi một máy trong một tuần có trung bình 20,9mm và độ lệch tiêu chuẩn 1,07mm. Ước lượng trung bình đường kính của viên bi với độ tin cậy (a) 95%, (b) 99%.

5. Để khảo sát sức bền chịu lực của một loại ống công nghiệp người ta tiến hành đo 9 ống và thu được các số liệu sau

   4500 6500 5000 5200 4800 4900 5125 6200 5375

   Từ kinh nghiệm nghề nghiệp người ta biết rằng sức bền đó có phân phối chuẩn với độ lệch chuẩn $\sigma = 300$. Xác định khoảng tin cậy 95% cho sức bền trung bình của loại ống trên.

6. Tại một vùng rừng nguyên sinh, người ta đeo vòng cho 1000 con chim. Sau một thời gian, bắt lại 200 con thì thấy có 40 con có đeo vòng. Thử ước lượng số chim trong vùng rừng đó với độ tin cậy 99%.

7. Biết tỷ lệ nảy mầm của một loại hạt giống là 0,9. Với độ tin cậy 0,95, nếu ta muốn độ dài khoảng tin cậy của tỷ lệ nảy mầm không vượt quá 0,02 thì cần phải gieo bao nhiêu hạt?

8. Kết quả quan sát về hàm lượng vitamine C của một loại trái cây cho ở bảng sau:

Hàm lượng vitamine C (%)	Số trái
6 − 7	5
7 − 8	10
8 − 9	20
9 − 10	35
10 − 11	25
11 − 12	5

   a) Hãy ước lượng hàm lượng vitamine C trung bình trong một trái với độ tin cậy 95%.
   b) Qui ước những trái có hàm lượng vitamine C trên 10% là trái loại A. Ước lượng tỷ lệ trái loại A với độ tin cậy 90%.
   c) Muốn độ chính xác khi ước lượng hàm lượng vitamine C trung bình là 0,1 và độ chính xác khi ước lượng tỷ lệ trái loại A là 5% với cùng độ tin cậy 95% thì cần quan sát thêm bao nhiêu trái nữa? A

9. Đo đường kính của 100 chi tiết máy do một phân xưởng sản xuất, ta được kết quả cho ở bảng sau:

Đường kính (mm)	Số chi tiết máy
9,85	8
9,90	12
9,95	20
10,00	30
10,05	14
10,10	10
10,15	6

Theo qui định, những chi tiết có đường kính tù $9,9mm$ đến $10,1mm$ là những chi tiết đạt tiêu chuẩn kỹ thuật.

a) Ước lượng tỷ lệ và ước lượng trung bình đường kính của những chi tiết đạt tiêu chuẩn với cùng độ tin cậy 95%?
b) Để độ chính xác khi ước lượng đường kính trung bình của những chi tiết đạt tiêu chuẩn là $0,02mm$ và độ chính xác khi ước lượng tỷ lệ chi tiết đạt tiêu chuẩn là 5% với cùng độ tin cậy 99% thì cần đo thêm ít nhất bao nhiêu chi tiết nữa?

10. Độ dài của bản kim loại tuân theo luật chuẩn. Đo 10 bản kim loại đó ta thu được số liệu sau:

$$4,1 \quad 3,9 \quad 4,7 \quad 4,4 \quad 4,0 \quad 3,8 \quad 4,4 \quad 4,2 \quad 4,4 \quad 5,0$$

Hãy xác định

a) Khoảng tin cậy 90% cho độ dài trung bình trên;
b) Khoảng tin câjy 95% cho phương sai của độ dài đó.

11. Người ta đo chiều sâu của biển, sai lệch ngẫu nhiên được giả thiết phân phối theo qui luật chuẩn với độ lệch tiêu chuẩn là $20m$. Cần đo bao nhiêu lần để xác định chiều sâu của biển với sai lệch không quá $15m$ và độ tin cậy đạt được 95%?

12. Theo dõi số hàng bán được trong một ngày ở một của hàng, ta được kết quả ghi ở bảng sau:

Số hàng bán được ($kg/ngày$)	Số ngày
1900 − 1950	2
1950 − 2000	10
2000 − 2050	8
2050 − 2100	5

Hãy ước lượng phương sai của lượng hàng bán được mỗi ngày với độ tin cậy 95%? (cho biết $\alpha_1 = \alpha_2$).

◙ TRẢ LỜI BÀI TẬP

1. a) $9,5kg$, b) $0,74kg^2$

2. $\overline{x} = 6,35cm$, $s^2 = 0,00055cm^2$.

3. a) Trung bình khối lượng $m = 95kg$. Ước lượng không chệch của phương sai là

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 = \frac{1}{5}\sum_{i=1}^{5}(x_i - 95)^2 = 0,41$$

b) $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \dfrac{1}{5}\sum_{i=1}^{5} x_i = 95,5$

Ước lượng không chệch của phương sai là

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{X})^2 = \frac{1}{4}\sum_{i=1}^{5}(x_i - 95,5)^2 = 0,7rff$$

4. (a) $20,9 \pm 0,148mm$, (b) $20,9 \pm 0,195mm$.

5. $(5092,89\,;\,5484,89)$.

6. $0,1271 < p < 0,2729$

Tổng số chim trong vùng rừng nằm trong khoảng $(\frac{1000}{0,2729}, \frac{1000}{0,1271})$

7. $2 \times 1,96\sqrt{\frac{0,9\times 0,1}{n}} < 0,02$. Giải bất phương trình ta có $n > 3457$.

8. a) $9,06; 9,54)$, c) 467 trái.

9. a) $(0,792 < p < 0,928)$; $(9,982 < m < 10,006)$. b) 221.

10. a) $(4,09\,;\,4,49)$, b) $(0,064\,;\,0,456)$.

11. 7 lần.

12. $(1253,8 < \sigma^2 < 3983,8)$.