CHƯƠNG VI: LÝ THUYẾT ƯỚC LƯỢNG

Sử dụng phương pháp quy nạp thống kê ta có thể ước lượng các tham số đặc trưng của tổng thể thông qua thống kê của mẫu.

Phương pháp ước lượng điểm chủ trương dùng giá trị quan sát của một thống kê để ước lượng một tham số (véc tơ tham số) nào đó theo các tiêu chuẩn: vững, không chệch, hiệu quả. Có nhiều phương pháp khác nhau để ước lượng điểm. Trong chương này chỉ xét phương pháp mômen và phương pháp hợp lý cực đại.

Khi kích thước mẫu bé thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng. Mặt khác phương pháp trên cũng không thể đánh giá được khả năng mắc sai lầm khi ước lượng là bao nhiêu. Do đó khi kích thước mẫu bé người ta thường dùng phương pháp ước lượng khoảng tin cậy cho trường hợp một tham số. Khoảng tin cậy là khoảng mà tham số của dấu hiệu nghiên cứu của tổng thể rơi vào khoảng này với xác suất bằng độ tin cậy.

Trong chương này ta sẽ xây dựng ước lượng ước lượng điểm và khoảng tin cậy cho kỳ vọng, phương sai của dấu hiệu nghiên cứu có phân bố chuẩn và tần suất của tổng thể.

6.1 PHƯƠNG PHÁP ƯỚC LƯỢNG ĐIỂM

6.1.1 Khái niệm ước lượng điểm

Phương pháp ước lượng điểm chủ trương dùng một giá trị để thay cho giá trị của tham số θ chưa biết của tổng thể. Thông thường giá trị được chọn này là giá trị cụ thể của một thống kê $\boldsymbol{\Theta}$ nào đó của mẫu ngẫu nhiên.

Với mẫu ngẫu nhiên $W=(X_1,X_2,...,X_n)$, thống kê ước lượng cho tham số θ được ký hiệu:

$$\boldsymbol{\Theta}=T\left(X_1,X_2,...,X_n\right)$$

Khi đó với mẫu cụ thể $w=(x_1,x_2,...,x_n)$ giá trị cụ thể của thống kê $\theta=T\left(x_1,x_2,...,x_n\right)$ là ước lượng của θ.

Cùng với một mẫu ngẫu nhiên có thể xây dựng nhiều thống kê $\boldsymbol{\Theta}$ khác nhau để ước lượng cho tham số θ. Vì vậy ta cần lựa chọn thống kê tốt nhất để ước lượng cho tham số θ dựa vào các tiêu chuẩn sau.

6.1.2 Một số loại ước lượng điểm

6..1.2.1 Ước lượng không chệch (unbiased estimator)

Thống kê $\boldsymbol{\Theta}=T\left(X_1,X_2,...,X_n\right)$ là một hàm của các biến ngẫu nhiên $X_1,X_2,...,X_n$ nên cũng là một biến ngẫu nhiên. Do đó ta có thể xét các đặc trưng của thống kê này.

Định nghĩa 6.1: *Thống kê* $\boldsymbol{\Theta} = T\left(X_1, X_2, ..., X_n\right)$ *được gọi là ước lượng không chệch của* θ *nếu với mọi giá trị của tham số* θ

$$\mathrm{E}\left[\boldsymbol{\Theta}\left(X_1, X_2, ..., X_n\right)\right] = \theta \qquad \textbf{(6.1)}$$

Nếu $\mathrm{E}(\boldsymbol{\Theta}) \neq \theta$ thì $\boldsymbol{\Theta} = T\left(X_1, X_2, ..., X_n\right)$ được gọi là một *ước lượng chệch* của θ.

6.1.2.2 Ước lượng hiệu quả (efficient estimator)

Điều kiện (6.1) của ước lượng không chệch có nghĩa rằng trung bình các giá trị của $\boldsymbol{\Theta}$ bằng giá trị θ. Tuy nhiên từng giá trị của $\boldsymbol{\Theta}$ có thể sai lệch rất lớn so với θ. Vì vậy ta tìm ước lượng không chệch sao cho độ sai lệch trung bình là bé nhất.

Định nghĩa 6.2: *Ước lượng không chệch có phương sai nhỏ nhất so với mọi ước lượng không chệch khác được xây dựng trên cùng một mẫu ngẫu nhiên gọi là ước lượng hiệu quả (hay ước lượng phương sai bé nhất).*

Như vậy, để xét xem ước lượng không chệch $\boldsymbol{\Theta}$ có phải là ước lượng hiệu quả của θ hay không ta cần phải tìm một cận dưới của phương sai của các ước lượng không chệch và so sánh phương sai của $\boldsymbol{\Theta}$ với cận dưới này. Điều này được giải quyết bằng bất đẳng thức Cramer-Rao phát biểu như sau:

Cho mẫu ngẫu nhiên $W = \left(X_1, X_2, ..., X_n\right)$ được lấy từ tổng thể có dấu hiệu nghiên cứu là biến ngẫu nhiên X mà hàm mật độ xác suất (hay hàm khối lượng xác suất) $f(x, \theta)$ thỏa mãn một số điều kiện nhất định (thường được thỏa mãn trong thực tế, ít ra là các phân bố xác suất đã xét trong chương II) và $\boldsymbol{\Theta}$ là ước lượng không chệch bất kỳ của θ thì

$$\mathrm{D}\left(\boldsymbol{\Theta}\right) \geq \frac{1}{n\,\mathrm{E}\left(\dfrac{\partial\left(\ln f(X,\theta)\right)}{\partial\theta}\right)^2} \qquad \textbf{(6.2)}$$

Ví dụ 6.1: Dựa vào bất đẳng thức Cramer-Rao ta có thể chứng minh được rằng trung bình mẫu $\overline{X}$ là ước lượng hiệu quả của kỳ vọng μ của dấu hiệu nghiên cứu X của tổng thể có phân bố chuẩn $\mathbf{N}(\mu, \sigma^2)$.

Giải: Thật vậy theo công thức (5.8) ta có $\mathrm{D}\left(\overline{X}\right) = \dfrac{\sigma^2}{n}$. Mặt khác theo (2.37) hàm mật độ của X có dạng

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

$$\Rightarrow \ln f(x, \mu) = -\ln\left(\sigma\sqrt{2\pi}\right) - \frac{(x-\mu)^2}{2\sigma^2} \Rightarrow \frac{\partial \ln f(x,\mu)}{\partial\mu} = \frac{x-\mu}{\sigma^2}$$

Vậy $$nE\left(\frac{\partial(\ln f(X,\mu))}{\partial\mu}\right)^2 = nE\left(\frac{X-\mu}{\sigma^2}\right)^2 = \frac{n}{\sigma^4}E(X-\mu)^2 = \frac{n}{\sigma^4}D(X) = \frac{n}{\sigma^2}.$$

Như vậy $D\left(\overline{X}\right) = \frac{\sigma^2}{n}$ đạt giá trị cực tiểu của bất đẳng thức Cramer-Rao, do đó trung bình mẫu $\overline{X}$ là ước lượng hiệu quả của μ.

6.1.2.3 Ước lượng vững (consistent estimator)

Định nghĩa 6.3: *Thống kê* $\mathbf{\Theta} = T\left(X_1, X_2, ..., X_n\right)$ *được gọi là ước lượng vững của tham số* θ *của biến ngẫu nhiên gốc* X *nếu* $\mathbf{\Theta} = T\left(X_1, X_2, ..., X_n\right)$ *hội tụ theo xác suất đến* θ *khi* $n \to \infty$.

Nghĩa là với mọi $\varepsilon > 0$, luôn có

$$\lim_{n\to\infty} P\left\{\left|\mathbf{\Theta} - \theta\right| < \varepsilon\right\} = 1. \qquad \textbf{(6.3)}$$

Hoặc một cách tương đương

$$\lim_{n\to\infty} P\left\{\left|\mathbf{\Theta} - \theta\right| \geq \varepsilon\right\} = 0. \qquad \textbf{(6.4)}$$

Theo hệ quả 4.2 của luật số lớn Trêbưsep và công thức (4.9) chương IV, ta có trung bình mẫu $\overline{X}$ là ước lượng vững của kỳ vọng μ; S^2 và $\overline{S}^2$ là ước lượng vững của phương sai σ^2 của biến ngẫu nhiên gốc X của tổng thể.

Theo luật số lớn Bernoulli ta suy ra tần suất mẫu f là ước lượng vững của xác suất p của tổng thể.

Tóm lại ta có kết quả sau:

- Trung bình mẫu $\overline{X}$ là ước lượng không chệch, hiệu quả và vững của kỳ vọng μ của biến ngẫu nhiên gốc của tổng thể.
- Tần suất mẫu f là ước lượng không chệch, hiệu quả và vững của xác suất p của tổng thể.
- Phương sai mẫu S^2 và S^{*2} (trường hợp μ đã biết) là ước lượng không chệch và vững của phương sai σ^2 của biến ngẫu nhiên gốc của tổng thể.

6.1.3 Một số phương pháp tìm ước lượng điểm

6.1.3.1 Phương pháp hợp lý cực đại (maximum-likelihood estimation)

Giả sử $\left(X_1, X_2, ..., X_n\right)$ là mẫu ngẫu nhiên của biến ngẫu nhiên gốc X, do đó $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập và có cùng phân bố với X.

Nếu X rời rạc với hàm khối lượng xác suất phụ thuộc tham số θ có dạng $p_X(x_k;\theta)$ thì hàm khối lượng xác suất đồng thời của mẫu $\left(X_1, X_2, ..., X_n\right)$ xác định như sau

$$p_{X_1...X_n}(x_{k_1},...,x_{k_n};\theta) = p_X(x_{k_1};\theta)...p_X(x_{k_n};\theta) \tag{6.5}$$

Nếu X liên tục có hàm mật độ xác suất phụ thuộc tham số θ có dạng $f_X(x;\theta)$ thì hàm mật độ xác suất đồng thời của mẫu $\left(X_1, X_2, ..., X_n\right)$:

$$f_{X_1...X_n}(x_1,...,x_n;\theta) = f_X(x_1;\theta)...f_X(x_n;\theta) \tag{6.6}$$

Đặt

$$L(\theta) = \begin{cases} p_{X_1...X_n}(x_{k_1},...,x_{k_n};\theta) & \text{nÕu } X \text{ rêi r¹c} \\ f_{X_1...X_n}(x_1,...,x_n;\theta) & \text{nÕu } X \text{ liªn tôc} \end{cases} \tag{6.7}$$

Khi θ là giá trị đúng của tham số thì hàm $L(\theta)$ biểu diễn tính hợp lý về các giá trị $x_1,...,x_n$ sẽ quan sát được. Vì vậy $L(\theta)$ (với θ là biến và $x_1,...,x_n$ là các tham số) được gọi là hàm hợp lý (likelihood function) của mẫu ngẫu nhiên.

Phương pháp hợp lý cực đại chủ trương rằng $\theta = g(x_1, x_2, ..., x_n)$ ước lượng tốt nhất cho tham số θ nếu: Khi ta xem $x_1,...,x_n$ là tham số còn θ là biến thì hàm hợp lý $L(x_1,...,x_n,\theta)$ đạt cực đại tại $\theta = g(x_1, x_2, ..., x_n)$.

Thống kê $\mathbf{\Theta} = g(X_1, X_2, ..., X_n)$ được gọi là ước lượng hợp lý cực đại của θ.

Nói cách khác ước lượng hợp lý cực đại là một thống kê phản ánh nhiều thông tin nhất về tham số θ cần ược lượng.

Giả sử $\theta = g(x_1,...,x_n)$ là giá trị cực đại của $L(\theta)$, nghĩa là

$$L(\theta) = \max_{\theta} L(\theta) \tag{6.8}$$

Thì ước lượng hợp lý cực đại của θ là

$$\mathbf{\Theta} = g(X_1,...,X_n) \tag{6.9}$$

Hàm hợp lý $L(\theta)$ là hàm khối lượng xác suất đồng thời hoặc hàm mật độ xác suất đồng thời do đó luôn luôn nhận giá trị dương, ngoài ra hàm lôgarít là hàm đồng biến nên $L(\theta)$ và $\ln L(\theta)$ đạt cực đại tại cùng một giá trị. Vì vậy có thể tìm điểm cực đại θ bằng cách tìm điểm cực đại của hàm $\ln L(\theta)$.

Hàm hợp lý đạt cực đại tại các điểm dừng θ (nếu hàm hợp lý khả vi thì chỉ có thể đạt cực đại tại những điểm dừng):

- Nếu $L(\theta)$ là hàm một biến số θ thì θ là nghiệm của phương trình

$$\frac{d}{d\theta}\ln L(\theta)=0 \qquad \textbf{(6.10)}$$

- Nếu $L(\theta)$ là hàm nhiều biến số $\theta=(\theta_1,...,\theta_n)$ thì $\theta=(\theta_1,...,\theta_n)$ là nghiệm của hệ phương trình

$$\begin{cases}\dfrac{\partial}{\partial\theta_1}\ln L(\theta_1,...,\theta_n)=0\\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\\ \dfrac{\partial}{\partial\theta_n}\ln L(\theta_1,...,\theta_n)=0\end{cases} \qquad \textbf{(6.11)}$$

Phương pháp này do R. A. Fisher đề xuất và là một phương pháp quan trọng để tìm thống kê ước lượng.

Ví dụ 6.2: Giả sử X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, trong đó σ^2 đã biết. Hãy tìm ước lượng hợp lý cực đại của μ.

Trong trường hợp này tham số θ là tham số μ, theo công thức (6.6), (6.7) ta có:

$$L(\theta)=L(\mu)=\frac{1}{(\sqrt{2\pi}\sigma)^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2}$$

Do đó $$\ln L(\theta)=\ln L(\mu)=n\ln\frac{1}{\sqrt{2\pi}\sigma}-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2 .$$

$$\frac{d}{d\theta}\ln L(\theta)=\frac{d}{d\mu}\ln L(\mu)=\frac{1}{\sigma^2}\sum_{i=1}^n(x_i-\mu) .$$

Giải phương trình (6.10) ta được nghiệm $\mu=\frac{1}{n}\sum_{i=1}^n x_i=\overline{x}$.

Ngoài ra $\frac{d^2}{d\mu^2}\ln L(\mu)=-\frac{n}{\sigma^2}<0$ nên $\ln L(\mu)$ đạt cực đại tại $\mu=\frac{1}{n}\sum_{i=1}^n x_i=\overline{x}$, nghĩa là thỏa mãn điều kiện (6.8).

Vậy ước lượng hợp lý cực đại của μ là

$$\boldsymbol{\mu}=\frac{1}{n}\sum_{i=1}^n X_i=\overline{X} .$$

Ví dụ 6.3: Giả sử X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, trong đó μ và σ^2 chưa biết. Hãy tìm ước lượng hợp lý cực đại đồng thời của hai tham số μ và σ^2.

Trong trường hợp này tham số θ là $(\mu;\sigma^2)$, theo công thức (6.6), (6.7) ta có:

$$L(\theta)=L(\mu;\sigma^2)=\frac{1}{(\sqrt{2\pi}\sigma)^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$\ln L(\theta)=\ln L(\mu;\sigma^2)=n\ln\frac{1}{\sqrt{2\pi\sigma^2}}-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2=-\frac{n}{2}\ln\sigma^2+n\ln\frac{1}{\sqrt{2\pi}}-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 .$$

$$\begin{cases}\dfrac{\partial}{\partial\mu}\ln L(\mu;\sigma^2)=\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)\\ \dfrac{\partial}{\partial\sigma^2}\ln L(\mu;\sigma^2)=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2(\sigma^2)^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2\end{cases}.$$

Giải hệ phương trình (6.11) ta được

$$\begin{cases}\mu=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_i=\overline{x}\\ \sigma^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\overline{x})^2=\hat{s}^2\end{cases}$$

Xét các đạo hàm riêng cấp 2 tại μ, σ^2:

$$A=\frac{\partial^2}{\partial\mu^2}\ln L(\mu;\sigma^2)=\frac{-n}{\sigma^2};\ C=\frac{\partial^2}{\partial(\sigma^2)^2}\ln L(\mu;\sigma^2)=\frac{n}{2\left(\sigma^2\right)^2}-\frac{1}{\left(\sigma^2\right)^3}\sum_{i=1}^{n}(x_i-\mu)^2=-\frac{n}{2\left(\sigma^2\right)^2}$$

$$B=\frac{\partial^2}{\partial\sigma^2\partial\mu}\ln L(\mu;\sigma^2)=-\frac{1}{\left(\sigma^2\right)^2}\sum_{i=1}^{n}(x_i-\mu)=-\frac{1}{\left(\sigma^2\right)^2}\left\{\left(\sum_{i=1}^{n}x_i\right)-n\mu\right\}=0 .$$

$$A<0 \text{ và } AC-B^2=\left(-\frac{n}{\sigma^2}\right)\left(-\frac{n}{2\left(\sigma^2\right)^2}\right)-0=\frac{n^2}{2\left(\sigma^2\right)^3}>0 .$$

Do đó $L(\theta)=L(\mu;\sigma^2)$ đạt cực đại tại μ, σ^2 .

Vậy ta nhận được ước lượng hợp lý cực đại của μ và σ^2 là

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i;\ \ S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2 .$$

6.1.3.2 Phương pháp mô men (moment estimation)

Cho mẫu ngẫu nhiên $\left(X_1,X_2,...,X_n\right)$ của biến ngẫu nhiên gốc X . Giả sử phân bố của X phụ thuộc vào m tham số $\left(\theta_1,\theta_2,...,\theta_m\right)$ và ta cần ước lượng m tham số này. Tính các mô men của biến ngẫu nhiên gốc ta được hệ

$$\begin{cases} \mathrm{E}\,X = g_1(\theta_1,\theta_2,...,\theta_m) \\ \mathrm{E}\,X^2 = g_2(\theta_1,\theta_2,...,\theta_m) \\ .... \\ \mathrm{E}\,X^m = g_m(\theta_1,\theta_2,...,\theta_m) \end{cases} \quad \textbf{(6.12)}$$

Thay mô men mẫu vào các vế trái của công thức (6.12) ta được

$$\begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} X_i = g_1(\theta_1,\theta_2,...,\theta_m) \\ \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 = g_2(\theta_1,\theta_2,...,\theta_m) \\ .... \\ \dfrac{1}{n}\sum_{i=1}^{n} X_i^m = g_m(\theta_1,\theta_2,...,\theta_m) \end{cases} \quad \textbf{(6.13)}$$

Giải hệ phương trình (6.13) ta được các thống kê $(\Theta_1,\Theta_2,...,\Theta_m)$ của $(X_1,X_2,...,X_n)$.

$(\Theta_1,\Theta_2,...,\Theta_m)$ là ước lượng của $(\theta_1,\theta_2,...,\theta_m)$ theo phương pháp mô men.

Ví dụ 6.4: Giả sử X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$; μ,σ^2 đều chưa biết biết. Hãy tìm ước lượng của μ và σ^2 theo phương pháp mô men.

Hệ phương trình (6.12) có dạng

$$\begin{cases} \mathrm{E}\,X = \mu \\ \mathrm{E}\,X^2 = \sigma^2 + \mu^2 \end{cases}$$

Suy ra hệ phương trình (6.13) có dạng

$$\begin{cases} \dfrac{1}{n}\sum_{i=1}^{n} X_i = \mu \\ \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 = \sigma^2 + \mu^2 \end{cases}$$

Suy ra $$\mu = \overline{X}\ ;\ \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mu^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = S^2.$$

Trong trường hợp này ước lượng tìm bằng phương pháp mô men trùng với ước lượng tìm bằng phương pháp hợp lý cực đại.

Ngoài hai phương pháp ước lượng hợp lý cực đại và phương pháp mô men còn có một số phương pháp khác như: phương pháp Bayes, phương pháp minimax, phương pháp bootstrap … Tuy nhiên ta không có điều kiện để xét các phương pháp này.

6.2 PHƯƠNG PHÁP ƯỚC LƯỢNG BẰNG KHOẢNG TIN CẬY

Phương pháp ước lượng điểm nói trên có nhược điểm là khi kích thước mẫu bé thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng. Mặt khác phương pháp trên cũng không thể đánh giá được khả năng mắc sai lầm khi ước lượng là bao nhiêu. Do đó khi kích thước mẫu bé người ta thường dùng phương pháp ước lượng khoảng tin cậy. Theo phương pháp này từ mẫu ngẫu nhiên ta có thể tìm được khoảng [a; b] chứa tham số θ với xác suất β đủ lớn cho trước (β được gọi là độ tin cậy và thường được chọn trong khoảng 0,95 đến 0,99).

Định nghĩa 6.4: *Khoảng $[a;b]$ có hai đầu mút là hai thống kê*

$$a = a(X_1, X_2, ..., X_n),\ b = b(X_1, X_2, ..., X_n) \quad \textbf{(6.14)}$$

phụ thuộc mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$ của biến ngẫu nhiên gốc X, gọi là khoảng tin cậy của tham số θ với độ tin cậy β nếu:

$$P\{a(X_1, X_2, ..., X_n) \le \theta \le b(X_1, X_2, ..., X_n)\} = \beta \quad \textbf{(6.15)}$$

Trong thực tế thường yêu cầu độ tin cậy β khá lớn, khi đó theo nguyên lý xác suất lớn biến cố $\{a \le \theta \le b\}$ hầu như chắc chắn sẽ xảy ra trong một phép thử.

Tiến hành một phép thử với mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$ ta thu được một mẫu cụ thể $w = (x_1, x_2, ..., x_n)$, tính được giá trị cụ thể $a = a(x_1, x_2, ..., x_n)$, $b = b(x_1, x_2, ..., x_n)$.

Lúc đó có thể kết luận là: Qua mẫu cụ thể với độ tin cậy β tham số θ của biến ngẫu nhiên gốc X sẽ nằm trong khoảng $[a;b]$, tức là $a \le \theta \le b$.

6.2.1 Khoảng tin cậy của kỳ vọng của biến ngẫu nhiên phân bố chuẩn

Giả sử biến ngẫu nhiên gốc X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ nhưng chưa biết tham số μ của nó.

Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n: $W = (X_1, X_2, ..., X_n)$, ta tìm khoảng tin cậy của μ trong các trường hợp sau:

6.2.1.1 Trường hợp phương sai σ^2 đã biết

Định lý 6.1: Khoảng tin cậy của tham số μ với độ tin cậy β có dạng:

$$\left[\overline{X} - U_{\alpha/2}\frac{\sigma}{\sqrt{n}};\ \overline{X} + U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \quad \textbf{(6.16)}$$

trong đó: $\alpha = 1 - \beta$; $U_{\alpha/2}$ là giá trị tới hạn mức $\alpha/2$ của phân bố chuẩn tắc $\mathbf{N}(0;1)$ (công thức 2.42).

Chứng minh: Theo công thức (5.22) ta có $\dfrac{(\overline{X} - \mu)\sqrt{n}}{\sigma} \sim \mathbf{N}(0;1)$.

Mặt khác: $\overline{X}-U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\le\mu\le\overline{X}+U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}} \Leftrightarrow \left|\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\right|\le U_{\alpha/2}$.

Áp dụng công thức (2.43) ta có

$$P\left[\overline{X}-U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\le\mu\le\overline{X}+U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]=P\left\{\left|\frac{(\overline{X}-\mu)\sqrt{n}}{\sigma}\right|\le U_{\alpha/2}\right\}=1-\alpha=\beta.$$

Định nghĩa 6.5: $\varepsilon=U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ được gọi là *độ chính xác của ước lượng*.

Với phương sai đã biết σ^2 không đổi và độ tin cậy β không đổi thì giá trị tới hạn $U_{\alpha/2}$ không đổi, do đó độ chính xác chỉ phụ thuộc vào kích thước mẫu n. Khi n càng lớn thì ε càng bé, do đó khoảng ước lượng càng chính xác. Nói cách khác độ chính xác phụ thuộc kích thước mẫu. Nếu muốn ước lượng với độ chính xác ε_0 và độ tin cậy β cho trước, *kích thước mẫu cần thiết* là số tự nhiên n nhỏ nhất thỏa mãn:

$$n\ge\frac{\sigma^2U_{\alpha/2}^2}{\varepsilon_0^2} \tag{6.17}$$

Ví dụ 6.5: Trọng lượng của một loại sản phẩm là một biến ngẫu nhiên phân bố chuẩn với độ lệch tiêu chuẩn 1 gram. Cần thử 25 sản phẩm loại này ta thu được kết quả:

Trọng lượng (gram)	18	19	20	21
Số SP tương ứng	3	5	15	2

Với độ tin cậy 95%

a) Hãy tìm khoảng tin cậy của trọng lượng trung bình của loại sản phẩm trên.

b) Nếu muốn độ chính xác của ước lượng không vượt quá 0,3 gram thì cần cân thử ít nhất bao nhiêu sản phẩm.

Giải: Gọi X là trọng lượng sản phẩm, theo giả thiết X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ với $\sigma=1$. Trọng lượng trung bình của sản phẩm là tham số μ. Khoảng tin cậy có dạng (6.16).

Với độ tin cậy $\beta=0,95\Rightarrow\dfrac{\alpha}{2}=0,025\Rightarrow U_{\frac{\alpha}{2}}=1,96$.

a) Từ bảng số liệu tìm được trung bình mẫu cụ thể:

$$\overline{x}=\frac{3\cdot18+5\cdot19+15\cdot20+2\cdot21}{25}=19,64.$$

Độ chính xác của ước lượng $\varepsilon=U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}=1,96\cdot\dfrac{1}{\sqrt{25}}=0,392$.

Vậy với độ tin cậy 95% qua mẫu cụ thể này, khoảng tin cậy của tham số μ là:

$$\left[19{,}64-0{,}392\,;\,19{,}64+0{,}392\right]$$

hay $$19,248 \le \mu \le 20,032.$$

b) Nếu muốn độ chính xác của ước lượng không vượt quá 0,3 thì cần cân thử ít nhất n sản phẩm sao cho:

$$n \ge \frac{\sigma^2 U_{\alpha/2}^2}{\varepsilon_0^2} = \frac{1\cdot 1,96^2}{0,3^2} = 42,68. \text{ Chọn } n = 43.$$

6.2.1.2 Trường hợp phương sai σ^2 chưa biết, kích thước mẫu $n \ge 30$

Trong nhiều bài toán thực tế, ta không biết phương sai σ^2 của biến ngẫu nhiên gốc X của tổng thể. Nhưng nếu kích thước mẫu n đủ lớn ($n \ge 30$) ta có thể xấp xỉ độ lệch chuẩn σ bởi độ lệch chuẩn mẫu S (vì S^2 là ước lượng vững không chệch của σ^2), S được xác định bởi công thức (5.13). Mặt khác, theo định lý giới hạn trung tâm thì thống kê $\dfrac{(\overline{X}-\mu)\sqrt{n}}{\sigma}$ có phân bố xấp xỉ phân bố chuẩn, đúng với mọi biến ngẫu nhiên gốc X (không đòi hỏi phân bố chuẩn).

Do đó khoảng tin cậy của tham số μ với độ tin cậy β có thể lấy là

$$\left[\overline{X} - U_{\alpha/2}\frac{S}{\sqrt{n}}\,;\,\overline{X} + U_{\alpha/2}\frac{S}{\sqrt{n}}\right] \qquad \textbf{(6.18)}$$

Ví dụ 6.6: Để xác định chiều cao trung bình của các cây bạch đàn trong khu rừng rộng trồng bạch đàn, ta tiến hành đo ngẫu nhiên 35 cây và có kết quả cho trong bảng sau:

Khoảng	r_i	x_i	$u_i = x_i - 8,25$	$r_i u_i$	$r_i u_i^2$
$6,5-7,0$	2	6,75	$-1,5$	-3	$4,5$
$7,0-7,5$	4	7,25	$-1,0$	-4	4
$7,5-8,0$	10	7,75	$-0,5$	-5	$2,5$
$8,0-8,5$	11	8,25	0	0	0
$8,5-9,0$	5	8,75	$0,5$	$2,5$	$1,25$
$9,0-9,5$	3	9,25	$1,0$	3	3
$\sum$	35			$-6,5$	$15,25$

$$\overline{u} = \frac{-6,5}{35} = -0,1857 \;\Rightarrow\; \overline{x} = 8,25 - 0,1857 \approx 8,06.$$

$$s^2 = s_u^2 = \frac{1}{34}\left(15,25 - \frac{(-6,5)^2}{35}\right) = 0,413 \;\Rightarrow\; s = 0,64.$$

Với độ tin cậy $\beta = 95\%$, $U_{\alpha/2} = 1,96$.

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2}\dfrac{s}{\sqrt{n}} = 1,96 \cdot \dfrac{0,64}{\sqrt{35}} = 0,21$.

Vậy khoảng tin cậy cho chiều cao trung bình μ của các cây bạch đàn là:

$$7,85 \le \mu \le 8,27.$$

6.2.1.3 Trường hợp phương sai σ^2 chưa biết, kích thước mẫu $n < 30$

Trong trường hợp này, theo công thức (5.25) thống kê

$$T = \frac{(\overline{X} - \mu)\sqrt{n}}{S} \quad \textbf{(6.19)}$$

có phân bố Student $n-1$ bậc tự do. Vì vậy khoảng tin cậy được tính theo kết quả sau:

Định lý 6.2: Khoảng tin cậy của tham số μ với độ tin cậy β có dạng:

$$\left[\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}};\ \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\right], \quad \textbf{(6.20)}$$

trong đó $t_{\alpha/2}(n-1)$ là giá trị tới hạn mức $\alpha/2$ của phân bố Student $n-1$ bậc tự do (công thức 2.57).

Độ chính xác của ước lượng:

$$\varepsilon = t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} \quad \textbf{(6.21)}$$

Với độ tin cậy β và độ chính xác ε_0 cho trước thì kích thước mẫu cần thiết là số tự nhiên n nhỏ nhất thỏa mãn:

$$n \ge \left(\frac{S \cdot t_{\alpha/2}(n-1)}{\varepsilon_0}\right)^2 \quad \textbf{(6.22)}$$

Ví dụ 6.7: Năng suất của một loại giống mới là một biến ngẫu nhiên có quy luật phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$. Gieo thử giống này trên 16 mảnh vườn thí nghiệm thu được như sau (đơn vị kg/ha):

172, 173, 173,174, 174, 175, 176, 166, 166, 167, 165, 173, 171, 170, 171, 170.

Hãy tìm khoảng tin cậy cho năng suất trung bình của loại giống này với độ tin cậy $\beta = 95\%$.

Giải: Năng suất trung bình của hạt giống là tham số μ.

Từ các số liệu trên ta tính được: $\overline{x} = 171;\ s = 3,4254$. $\alpha = 0,05;\ \dfrac{\alpha}{2} = 0,025$.

Tra bảng phân bố Student với 15 bậc tự do ta tìm được $t_{\alpha/2}(n-1) = t_{0,025}(15) = 2,131$.

Độ chính xác $\varepsilon = t_{\alpha/2}(n-1)\dfrac{S}{\sqrt{n}} = 2,131\cdot\dfrac{3,4254}{\sqrt{16}} = 1,885$.

Vậy khoảng tin cậy cho năng suất trung bình của loại hạt giống này là μ thỏa mãn:

$$169,115 \le \mu \le 172,885.$$

6.2.2 Khoảng tin cậy cho xác suất p

Ta cần nghiên cứu một dấu hiệu định tính A nào đó mà mỗi cá thể của tổng thể có thể có hoặc không. Nếu cá thể có dấu hiệu A ta cho nhận giá trị 1, trường hợp ngược lại ta cho nhận giá trị 0. Lúc đó dấu hiệu nghiên cứu có thể xem là biến ngẫu nhiên X có phân bố Bernoulli với tham số p. Kỳ vọng $\mathrm{E}X = p$ và phương sai $\mathrm{D}X = p(1-p)$.

Lấy mẫu ngẫu nhiên $W = \left(X_1, X_2, ..., X_n\right)$, trong đó $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố Bernoulli tham số p.

Tần suất mẫu $f = \dfrac{X_1 + \cdots + X_n}{n}$.

Theo định lý Moivre-Laplace và công thức (4.12) ta có thể xấp xỉ thống kê $U = \dfrac{(f-p)\sqrt{n}}{\sqrt{p(1-p)}}$ với phân bố chuẩn tắc $\mathbf{N}(0;1)$ khi n đủ lớn.

Tuy nhiên vì p chưa biết nên chưa biết $p(1-p) = \mathrm{D}X$.

Mặt khác theo công thức (5.15), (5.26), (5.27) thì tần suất mẫu f là ước lượng vững, không chệch và hiệu quả của xác suất p tổng thể. Vì vậy khi n đủ lớn ta có thể thay p bằng f.

Do đó khoảng tin cậy cho xác suất p của tổng thể với độ tin cậy β là:

$$\left[f - U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}\ ;\ f + U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \right] \tag{6.23}$$

Với điều kiện n đủ lớn thỏa mãn

$$\begin{cases} nf > 10 \\ n(1-f) > 10 \end{cases} \tag{6.24}$$

trong đó $U_{\alpha/2}$ là giá trị tới hạn mức $\dfrac{\alpha}{2}$ của phân bố chuẩn tắc $\mathbf{N}(0;1)$ với $\alpha = 1-\beta$.

Độ chính xác của khoảng tin cậy:

$$\varepsilon = U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}. \tag{6.25}$$

Với độ tin cậy β và độ chính xác ε_0 cho trước thì *kích thước mẫu cần thiết* là số tự nhiên n nhỏ nhất thỏa mãn:

$$n \geq f(1-f)\left(\frac{U_{\alpha/2}}{\varepsilon_0}\right)^2 \quad \textbf{(6.26)}$$

trong đó f là tần suất mẫu của một mẫu ngẫu nhiên nào đó.

Ví dụ 6.8: Trong đợt vận động bầu cử tổng thống ở một nước nọ, người ta phỏng vấn ngẫu nhiên 1600 cử tri, được biết có 960 người trong số đó sẽ bỏ phiếu cho ứng cử viên A. Với độ tin cậy $\beta = 95\%$ thì ứng cử viên A sẽ chiếm được tối thiểu bao nhiêu % số phiếu bầu.

Giải: Gọi p là tỉ lệ số phiếu sẽ bầu cho ứng cử viên A. Tổng thể nghiên cứu là tập hợp tất cả các cử tri. Dấu hiệu nghiên cứu là cử tri sẽ bỏ phiếu cho A, có thể xem là biến ngẫu nhiên có phân bố Bernoulli tham số p. Khoảng tin cậy cho p có dạng (6.23) với điều kiện (6.24).

Từ mẫu cụ thể trên ta có $f = \dfrac{960}{1600} = 0,6$ thỏa mãn điều kiện $\begin{cases} nf = 960 > 10 \\ n(1-f) = 640 > 10 \end{cases}$.

Độ chính xác của ước lượng $\varepsilon = U_{\alpha/2}\sqrt{\dfrac{f(1-f)}{n}} = 1,96\sqrt{\dfrac{0,6 \cdot 0,4}{1600}} = 0,024$.

Khoảng tin cậy: $0,576 \leq p \leq 0,624$.

Vậy với độ tin cậy 95% thì tối thiểu có 57,6% cử tri sẽ bỏ phiếu cho ứng cử viên A.

6.2.3 Ước lượng phương sai của biến ngẫu nhiên có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$

Giả sử trong tổng thể biến ngẫu nhiên gốc X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ nhưng chưa biết phương sai σ^2 của nó.

Từ tổng thể rút ra một mẫu ngẫu nhiên kích thước n: $W = (X_1, X_2, ..., X_n)$.

Ta sẽ chọn thống kê thích hợp để ước lượng cho tham số σ^2 trong các trường hợp sau:

6.2.3.1. Đã biết kỳ vọng μ

Chọn thống kê

$$T = \frac{nS^{*2}}{\sigma^2} \quad \textbf{(6.27)}$$

Theo công thức (5.23) thống kê T có phân bố khi bình phương n bậc tự do: $\chi^2(n)$. Do đó với độ tin cậy β cho trước, với cặp số α_1, α_2 sao cho $\alpha_1 + \alpha_2 = 1 - \beta$ có thể tìm hai giá trị tới hạn của T mức α_1, α_2 là $\chi^2_{1-\alpha_1}(n)$, $\chi^2_{\alpha_2}(n)$:

$$P\left\{T > \chi^2_{1-\alpha_1}(n)\right\} = 1 - \alpha_1 \text{ và } P\left\{T > \chi^2_{\alpha_2}(n)\right\} = \alpha_2 .$$

Do đó

$$P\left\{\chi^2_{1-\alpha_1}(n) < T < \chi^2_{\alpha_2}(n)\right\} = 1-(\alpha_1+\alpha_2) = \beta. \quad \textbf{(6.28)}$$

Thay thống kê T từ công thức (6.27) vào (6.28) và giải theo σ^2, ta được:

$$P\left\{\frac{nS^{*2}}{\chi^2_{\alpha_2}(n)} < \sigma^2 < \frac{nS^{*2}}{\chi^2_{1-\alpha_1}(n)}\right\} = \beta. \quad \textbf{(6.29)}$$

Như vậy, với độ tin cậy β khoảng tin cậy của σ^2 có dạng:

$$\left(\frac{nS^{*2}}{\chi^2_{\alpha_2}(n)}; \frac{nS^{*2}}{\chi^2_{1-\alpha_1}(n)}\right) \quad \textbf{(6.30)}$$

- Nếu $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$ khoảng tin cậy đối xứng có dạng:

$$\left(\frac{nS^{*2}}{\chi^2_{\alpha/2}(n)}; \frac{nS^{*2}}{\chi^2_{1-\alpha/2}(n)}\right) \quad \textbf{(6.31)}$$

- Nếu $\alpha_1 = 0;\ \alpha_2 = \alpha$ khoảng tin cậy bên phải của σ^2 có dạng:

$$\left(\frac{nS^{*2}}{\chi^2_{\alpha}(n)}; +\infty\right) \quad \textbf{(6.32)}$$

- Nếu $\alpha_2 = 0; \alpha_1 = \alpha$ khoảng tin cậy bên trái của σ^2 có dạng:

$$\left(0; \frac{nS^{*2}}{\chi^2_{1-\alpha}(n)}\right) \quad \textbf{(6.33)}$$

Nếu không nói rõ tìm khoảng tin cậy bên phải hay bên trái ta thì ta quy ước là cần tìm khoảng tin cậy đối xứng.

Ví dụ 6.9: Mức hao phí nguyên liệu cho 1 đơn vị sản phẩm là biến ngẫu nhiên có phân bố chuẩn với trung bình là 20 gam. Để ước lượng mức độ phân tán của mức hao phí này người ta cân thử 25 sản phẩm và thu được kết quả sau:

Hao phí nguyên liệu (gam)	19,5	20,0	20,5
Số sản phẩm tương ứng	5	18	2

Với độ tin cậy $\beta = 90\%$ hãy tìm khoảng tin cậy đối xứng của σ^2 nếu $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2} = 0,05$.

Giải: Gọi X là mức hao phí nguyên liệu cho 1 đơn vị sản phẩm. X có phân bố chuẩn với kỳ vọng đã biết $\mu = 20$. Đây là ước lượng phương sai σ^2 của phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ khi đã biết μ. Khoảng tin cậy theo công thức (2.22).

Tra bảng $\chi^2(n)$ ta có:

$$\chi^2_{\alpha/2}(n) = \chi^2_{0,05}(25) = 37,625;\ \chi^2_{1-\alpha/2}(n) = \chi^2_{0,95}(25) = 14,611.$$

Để tìm s^{*2} ta lập bảng sau:

x_i	r_i	$x_i - \mu$	$(x_i-\mu)^2$	$r_i(x_i-\mu)^2$
19,5	5	$-0,5$	0,25	1,25
20,0	18	0,0	0,00	0,00
20,5	2	0,5	0,25	0,50
$\sum$	25			1,75

$$s^{*2} = \frac{1,75}{25} = 0,07 .$$

Vậy với độ tin cậy 90%, qua mẫu cụ thể này, khoảng tin cậy của σ^2 là:

$$\left(\frac{25\cdot 0,07}{37,625}; \frac{25\cdot 0,07}{14,611}\right) \text{ hay } \quad 0,0465 < \sigma^2 < 0,1198 .$$

6.2.3.2 Chưa biết kỳ vọng μ

Chọn thống kê

$$T = \frac{(n-1)S^2}{\sigma^2} \tag{6.34}$$

Theo công thức (5.24) thống kê T có phân bố khi bình phương $n-1$ bậc tự do: $\chi^2(n-1)$. Do đó với độ tin cậy β cho trước, với cặp số α_1, α_2 sao cho $\alpha_1+\alpha_2 = 1-\beta$ có thể tìm hai giá trị tới hạn của T mức α_1, α_2 là $\chi^2_{1-\alpha_1}(n-1)$, $\chi^2_{\alpha_2}(n-1)$:

$$P\left\{T > \chi^2_{1-\alpha_1}(n-1)\right\} = 1-\alpha_1 \quad \text{và} \quad P\left\{T > \chi^2_{\alpha_2}(n-1)\right\} = \alpha_2 .$$

Do đó

$$P\left\{\chi^2_{1-\alpha_1}(n-1) < T < \chi^2_{\alpha_2}(n-1)\right\} = 1-(\alpha_1+\alpha_2) = \beta . \tag{6.35}$$

Thay thống kê T từ công thức (6.34) vào (6.35) và giải theo σ^2, ta được:

$$P\left\{\frac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right\} = \beta . \tag{6.36}$$

Như vậy, với độ tin cậy β khoảng tin cậy của σ^2 có dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha_2}(n-1)}\,;\frac{(n-1)S^2}{\chi^2_{1-\alpha_1}(n-1)}\right) \tag{6.37}$$

• Nếu $\alpha_1 = \alpha_2 = \dfrac{\alpha}{2}$ khoảng tin cậy đối xứng có dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\,;\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right) \tag{6.38}$$

• Nếu $\alpha_1 = 0; \alpha_2 = \alpha$ khoảng tin cậy bên phải của σ^2 có dạng:

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha}(n-1)}\,;+\infty\right) \tag{6.39}$$

• Nếu $\alpha_2 = 0; \alpha_1 = \alpha$ khoảng tin cậy bên trái của σ^2 có dạng:

$$\left(0\,;\frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}\right) \tag{6.40}$$

TÓM TẮT

Trung bình mẫu $\overline{X}$ là ước lượng không chệch, hiệu quả và vững của kỳ vọng μ của biến ngẫu nhiên gốc của tổng thể.

Tần suất mẫu f là ước lượng không chệch, hiệu quả và vững của xác suất p của tổng thể.

Phương sai mẫu S^2 và S^{*2} (trường hợp μ đã biết) là ước lượng không chệch và vững của phương sai σ^2 của biến ngẫu nhiên gốc của tổng thể.

Khoảng tin cậy của kỳ vọng của biến ngẫu nhiên gốc có phân bố chuẩn:

- Trường hợp phương sai σ^2 đã biết: công thức (6.16).
- Trường hợp phương sai σ^2 chưa biết, $n \ge 30$: công thức (6.18).
- Trường hợp phương sai σ^2 chưa biết, $n < 30$: công thức (6.20).

Khoảng tin cậy cho xác suất p: công thức (6.23).

Khoảng tin cậy của phương sai của biến ngẫu nhiên gốc có phân bố chuẩn:

- Trường hợp kỳ vọng μ đã biết: công thức (6.30)-(6.33).
- Trường hợp kỳ vọng μ chưa biết: công thức (6.37)-(6.40).

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

6.1 Trung bình mẫu là ước lượng vững và hiệu quả của kỳ vọng của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐.

6.2 Có thể tìm được ước lượng không chệch của θ có phương sai nhỏ hơn đại lượng $\dfrac{1}{n\mathrm{E}\left(\dfrac{\partial(\ln f(X,\theta))}{\partial\theta}\right)^2}$.

Đúng ☐ Sai ☐.

6.3 Tổng của hai ước lượng không chệch là một ước lượng không chệch.

Đúng ☐ Sai ☐.

6.4 Phương sai mẫu hiệu chỉnh S^2 là ước lượng vững không chệch của phương sai của biến ngẫu nhiên gốc.

Đúng ☐ Sai ☐.

6.5 Hai đầu mút của khoảng tin cậy là hai thống kê của mẫu.

Đúng ☐ Sai ☐.

6.6 Muốn tìm khoảng tin cậy cho tham số μ của biến ngẫu nhiên gốc có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì kích thước mẫu n phải lớn hơn 30.

Đúng ☐ Sai ☐.

6.7 Từ tổng thể có dấu hiệu nghiên cứu X có bảng phân bố xác suất sau

X	0	1
P	0,5	0,5

lập mẫu ngẫu nhiên kích thước $n=10$. Tính xác suất để trung bình mẫu của mẫu ngẫu nhiên này nhận giá trị 0,5.

6.8 Giả sử biến ngẫu nhiên gốc có phân bố chuẩn $\mathbf{N}(20;1)$. Chọn mẫu ngẫu nhiên kích thước $n=100$. Hãy tính xác suất để trung bình mẫu $\overline{X}$ nằm trong khoảng: $19{,}8<\overline{X}<20{,}2$.

6.9 Một mẫu cụ thể của biến ngẫu nhiên X như sau:

$$2\ ;\ 3\ ;\ 2\ ;\ 4\ ;\ 1\ ;\ 4\ ;\ 2\ ;\ 2\ ;\ 3\ ;\ 1\ \ (n=10).$$

a) Lập bảng phân bố tần suất.

b) Xây dựng hàm phân bố thực nghiệm. Tính $\overline{x}$, s^2, s.

6.10 Giả sử $\Theta_n=T(X_1,X_2,...,X_n)$ là một ước lượng của θ thỏa mãn:

$$\lim_{n\to\infty}\mathrm{E}\left(\Theta_n\right)=\theta \text{ và } \lim_{n\to\infty}\mathrm{D}\left(\Theta_n\right)=0.$$

Chứng minh Θ_n là một ước lượng vững của θ.

6.11 Giả sử $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên của biến ngẫu nhiên gốc X có phân bố Poisson tham số λ. Hãy ước lượng hợp lý cực đại tham số λ.

6.12 Giả sử $(X_1, X_2, ..., X_n)$ là một mẫu ngẫu nhiên của biến ngẫu nhiên gốc X có phân bố mũ tham số λ. Hãy ước lượng hợp lý cực đại tham số λ.

6.13 Trong đợt vận động bầu cử tổng thống ở một nước nọ, người ta phỏng vấn ngẫu nhiên 2000 cử tri thì được biết có 1082 người trong số đó sẽ bỏ phiếu cho ứng cử viên A. Với độ tin cậy 98% tối thiểu ứng cử viên A sẽ chiếm được bao nhiêu % số phiếu bầu?

6.14 Để xác định sản lượng khai thác điện thoại của đơn vị mình, một đơn vị đã tiến hành thống kê ngẫu nhiên 35 ngày và thu được kết quả sau với đơn vị 100.000 phút/ngày:

0,84 0,96 1,02 1,08 0,88 0,80 0,91 0,97 1,07 0,98 1,04 1,13 0,87 0,82 1,01
0,93 1,03 1,10 0,97 1,05 0,83 0,76 0,95 1,15 1,00 1,05 1,14 0,89 0,81
0,95 1,20 1,16 1,24 0,79 0,77.

Tìm khoảng tin cậy 95% cho sản lượng điện thoại trung bình mỗi ngày.

6.15 Muốn ước lượng số cá trong hồ, người ta bắt 2000 con cá trong hồ đánh dấu rồi thả lại xuống hồ. Sau đó bắt lại 400 con và thấy có 53 con có dấu. Hãy ước lượng số cá trong hồ với độ tin cậy là 0,95.

6.16 Để xác định chiều cao trung bình của các cây con trong một vườn ươm người ta tiến hành đo ngẫu nhiên 40 cây. Kết quả đo được như sau:

Khoảng chiều cao (cm)	16,5-17	17-17,5	17,5-18	18-18,5	18,5-19	19-19,5
Số cây tương ứng	3	5	11	12	6	3

a) Tìm khoảng tin cậy 90% cho chiều cao trung bình của vườn cây con.

b) Nếu muốn khoảng ước lượng có độ chính xác $\varepsilon = 0,1$ thì cần lấy mẫu bao nhiêu cây.

6.17 Trọng lượng của một loại sản phẩm A là một biến ngẫu nhiên có phân bố theo quy luật chuẩn với độ lệch chuẩn là 1 gam. Cân thử 27 bao loại này ta thu được kết quả:

Trọng lượng (gram)	47,5-48,5	48,5-49,5	49,5-50,5	50,5-51,5	51,5-52,5
Số bao tương ứng	3	6	15	2	1

a) Tìm khoảng tin cậy 95% của trọng lượng trung bình của loại sản phẩm trên.

b) Nếu muốn độ chính xác $\varepsilon = 0,1$ thì kích thước mẫu cần thiết là bao nhiêu.

6.18 Người ta đo một đại lượng không đổi 25 lần bằng một dụng cụ đo không có sai số hệ thống và sai số đo trung bình bằng 0. Giả sử sai số của phép đo tuân theo quy luật phân bố chuẩn và phương sai mẫu đo được bằng 0,5. Hãy xác định khoảng tin cậy 95% cho phương sai của sai số đo.