CHƯƠNG 7. ƯỚC LƯỢNG THAM SỐ

Bài 7.1. Ước lượng điểm cho kỳ vọng, median, phương sai và xác suất

1.Khái niệm về ước lượng điểm cho tham số:
Giả sử X là ĐLNN có tham số đặc trưng θ nào đó (chưa biết) mà ta đang quan tâm. Vấn đề đặt ra là: căn cứ trên n giá trị $x_1, x_2, \ldots, x_n$ của X được quan trắc trên một mẫu ngẫu nhiên (MNN) cỡ n lấy ra từ tập chính, cần đưa ra một giá trị gần đúng $\hat{\theta}$ của θ.

Định nghĩa 1. Một hàm $\hat{\theta} = \hat{\theta}_n(X_1, X_2, ..., X_n)$ của n giá trị $X_1, X_2, \ldots, X_n$ được gọi là một ước lượng điểm cho θ.
Để khảo sát về mặt toán học, ta coi $(x_1, x_2, \ldots, x_n)$ là giá trị quan sát được là một thể hiện của MNN cỡ n $(X_1, X_2, \ldots, X_n)$, trong đó $X_1, X_2, \ldots, X_n$ là các BNN độc lập cùng phân phối với X.

Như vậy, $\hat{\theta} = \hat{\theta}_n(X_1, X_2, ..., X_n)$ là 1 hàm của n BNN $X_1, X_2, \ldots, X_n$ và do đó cũng là 1 BNN. Giá trị của ước lượng cũng thay đổi từ mẫu quan sát này tới mẫu quan sát khác.
Việc lựa chọn một ước lượng nào là "tốt" được căn cứ trên các tiêu chuẩn sau:

2. Các tính chất của ước lượng điểm:

Định nghĩa 2.

1.Ước lượng $\hat{\theta}_n$ *gọi là ước lượng không chệch cho* θ *nếu* $E_\theta(\hat{\theta}_n) = \theta$.

Tính chất không chệch có nghĩa là ước lượng $\hat{\theta}_n$ không có sai số hệ thống.

2. *Ước lượng* $\hat{\theta}_n$ *gọi là ước lượng vững nếu với mọi* $\varepsilon > 0$

$$\lim_{n\to\infty} P\left\{\left|\hat{\theta}_n - \theta\right| < \varepsilon\right\} = 1.$$

Hay

$$\lim_{n\to\infty} P\left\{\theta - \varepsilon < \hat{\theta}_n < \theta + \varepsilon\right\} = 1$$

3. Ước lượng hiệu quả: Đó là ước lượng không chệch có phương sai nhỏ nhất trên lớp các ước lượng không chệch của θ.

3.Ước lượng điểm của giá trị trung bình:

Giả sử X là BNN với giá trị trung bình với E(X)= μ (chưa biết), μ được gọi là giá trị trung bình của tập hợp chính.

Ước lượng điểm của kỳ vọng là trung bình mẫu:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{\sum_{k=1}^{m} r_k x_k}{\sum_{k=1}^{m} r_k}$$

Định lý 1. Trung bình mẫu là ước lượng không chệch và vững cho giá trị trung bình μ của tập chính.

4.Ước lượng điểm của phương sai:
Ước lượng điểm của phương sai là phương sai mẫu:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\sum_{k=1}^{m} r_k (x_k - \overline{x})^2$$

Định lý 2. Phương sai mẫu là ước lượng không chệch và vững cho giá trị phương sai σ^2 của tập chính.

5. Ước lượng điểm của xác suất:

Giả sử A là biến cố mà ta quan tâm với p=P(A) chưa biết. Tiến hành quan sát n lần độc lập, ký hiệu m là tần số xuất hiện của A. Khi đó, $\hat{p} = m/n$ là ước lượng điểm của p.

Định lý 3. $\hat{p} = m/n$ là ước lượng này không chệch và vững cho giá trị xác suất p=P(A).

6. Ước lượng điểm của Median:

Ước lượng điểm của Median là Median mẫu được xác định như sau:
Với một mẫu, trung vị mẫu là là giá trị nằm giữa dãy giá trị quan trắc theo thứ tự tăng hay giảm.
Nếu dãy quan trắc có 2n+1 số liệu sắp xếp theo thứ tự tăng dần thì giá trị thứ n+1 là trung vị, nếu dãy quan trắc gồm 2n số liệu thì trung vị là giá trị trung bình của giá trị thứ n và n+1.
Nếu các giá trị x_i có tần số r_i, gọi k là chỉ số bé nhất để $r_1+r_2+\ldots+r_k \geq n/2$. Khi đó ta định nghĩa $Med(X)=x_k$.
Ví dụ: Cho bảng phân bố tần số của đại lượng X như sau:

x_i	0	1	2	3	4	5	6	7	8	9	10	11
r_i	6	15	43	53	85	72	55	33	18	10	7	3

Kích thước mẫu là 400
Hãy tính trung bình mẫu và trung vị.
Giải

Trung bình mẫu $\bar{x} = 4.645$
Ta thấy số giá trị của mẫu bé hơn hay bằng 3 là:
3+15+43+53=117<200
Số giá trị của mẫu bé hơn hay bằng 4 là:
3+15+43+53+85=202>200
Vậy Med(X)=4.
Trong trường hợp mẫu được cho dưới dạng phân bố ghép lớp ta định nghĩa trung vị như sau:
Giả sử ta có m khoảng với các điểm chia là:

$$a_0 < a_1 < \ldots < a_m$$

$C_1 = [a_0, a_1)$, $C_2 = [a_1, a_2)$, …, $C_m = [a_{m-1}, a_m]$. Trong đó khoảng C_i có tần số r_i.
Khoảng C_k được gọi là khoảng trung vị nếu k là chỉ số bé nhất sao cho $r_1+r_2+\ldots+r_k \geq n/2$.
Số trung vị Med(X) là số mà tại đó đường thẳng x=Med(X) chia đôi diện tích của tổ chức đồ tần số (tần suất).
$Med(X)=a_{k-1}+[(n/2)-(r_1+r_2+\ldots+r_{k-1})]/h_k$, h_k – là chiều cao của hình chữ nhật thứ k.

Bài 7.2. Ước lượng khoảng (Khoảng tin cậy)

Định nghĩa 2. Khoảng có 2 đầu mút $a(X_1, X_2, \ldots, X_n)$ và $b(X_1, X_2, \ldots, X_n)$ gọi là khoảng tin cậy với độ tin cậy $\gamma=1-\alpha$ của tham số θ, nếu:

$$P\{a \le \theta \le b\} = \gamma$$

1. Ước lượng khoảng cho giá trị trung bình:

A.Khi X là BNN chuẩn $N(\mu, \sigma^2)$:

A1. Ước lượng khoảng cho giá trị trung bình khi phương sai $\sigma^2 = \sigma_0^{\ 2}$ đã biết:

$$P\left\{\overline{x} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} \le \mu \le \overline{x} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right\} = 1-\alpha$$

Tra bảng Excel $z_{\alpha/2}$

=NORMSINV(1-α/2)

$z_{0.05}$=NORMSINV(1-0.05)

Do $\overline{X}$ có phân phối chuẩn $N(\mu, \frac{\sigma_0^2}{n})$.

Ví dụ 1. Hãy tìm khoảng tin cậy 95% cho chiều cao trung bình của sinh viên dựa trên một mẫu kích thước n=36 với trung bình mẫu $\overline{x} = 66\, inches$. Giả sử rằng độ lệch tiêu chuẩn σ_0 của chiều cao người lớn là 3 inches.

Giải

Ta có σ_0 =3, n=36, γ=95%, $z_{\alpha/2}$=1.96.

Vậy khoảng tin cậy 95% là :

$\bar{x} \pm z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} = 66 \pm 1.96\frac{3}{\sqrt{36}} = 66 \pm 1.96(0.5) = 66 \pm 0.98$ hay $[65.02; 66.98]$.

Vậy với độ tin cậy 95%, chiều cao trung bình μ nằm giữa 65.02 và 66.98 (inches).

Hoặc

Khoảng tin cậy 95% cho chiều cao trung bình μ là $[65.02; 66.98]$.

Ví dụ 2. Cũng câu hỏi như trên nhưng cần tìm khoảng tin cậy 99%.

Giải

Ta có σ=3, n=36, γ=99%, $z_{\alpha/2}$=2.58.

Vậy khoảng tin cậy 99% là :

$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 66 \pm 2.58\frac{3}{\sqrt{36}} = 66 \pm 2.58(0.5) = 66 \pm 1.29$ hay $[64.71; 67.29]$.

Vậy với độ tin cậy 99%, chiều cao trung bình μ nằm giữa 64.71 và 67.29 (inches).

A2. Ước lượng khoảng cho giá trị trung bình khi phương sai σ^2 chưa biết:

- Nếu n < 120 thì:

$$P\left\{\bar{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}$$

Trong đó $t_{\alpha/2}(n-1)$ tính theo phân phối Student với n-1 bậc tự do, tức là $P\{T > t_{\alpha/2}(n-1)\} = \alpha/2$ với T là BNN Student với n-1 bậc tự do.

Cơ sở cho việc xây dựng khoảng tin cậy trong trường hợp này là Mệnh đề sau:

Mệnh đề : Giả sử X_j với j=1, 2, ...n là các biến ngẫu nhiên Gauss độc lập cùng phân phối với X. Khi đó :

$$W = \frac{M_n - \mu}{V_n/\sqrt{n}} = \frac{\sqrt{n}(M_n - \mu)/\sigma}{V_n/\sigma} = \frac{(M_n - \mu)(\sigma/\sqrt{n})}{\left\{\left[(n-1)V_n^2/\sigma^2\right]/(n-1)\right\}^{1/2}}.$$

Có phân phối Student với (n-1) bậc tự do với hàm mật độ:

$$f_{n-1}(y) = \frac{\Gamma(n/2)}{\Gamma((n-1)/2)\sqrt{\pi(n-1)}}\left(1 + \frac{y^2}{n-1}\right)^{-n/2}$$

Ví dụ 3: Để xác định trọng lượng trung bình của các bao bột mỳ được đóng bao bằng máy tự động, người ta chọn ngẫu nhiên 15 bao và tính được $\overline{x} = 39.8$ và $S^2 = 0.144$.

Tìm khoảng tin cậy γ=99% cho trọng lượng trung bình μ của bao bột.

Giải

Ta có α=1-γ=1-0.99=0.01 ; $\frac{\alpha}{2} = 0.005$. Tra bảng phân phối Student với 14 bậc tư do ta tìm được t=TINV(0.01,14) = $t_{0.005}(14)=2.977$.

Vậy khoảng tin cậy 99% của μ là :

$$\bar{x} \pm t_{0.005}(14).(\frac{S}{\sqrt{N}}) = 39.8 \pm 2.977\left(\frac{0.3}{\sqrt{1}}\right.$$

Hay $39.5023 \leq \mu \leq 40.0977$

Ví dụ 4. Để ước lượng chiều cao trung bình μ của thanh niên của tỉnh Bắc Giang tham gia ngày Hội trái cây lớn nhất Miền Bắc ngày 25 – 26 tháng 5 năm 2017, một mẫu ngẫu nhiên gồm 16 thanh niên được chọn. Chiều cao của các thanh niên này đo được như sau (đơn vị cm) :

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

Hãy tìm khoảng tin cậy γ=99% cho chiều cao trung bình μ của thanh niên tỉnh Bắc Giang tham gia ngày Hội trái cây lớn nhất Miền Bắc ngày 25 – 26 tháng 5 năm 2017.

TB Mẫu	171.5625	t0.005(15)	2.947

PS mẫu	10.79583	γ=99%	
cận dưới	169.142		
cận trên	173.983		

Ví dụ ngày 30 tháng 7 năm 2019

n = 40

TBM	166.425
PSM	28.91731

Xây dựng khoảng tin cậy 95% cho chiều cao TB của SV Khoa CHKT&TĐH năm 2012. Biết rằng chiều cao của sv tuân theo luật chuẩn trong 2 trương hợp.
Trường hợp 1. Phương sai đã biết $\sigma_0^2 = 25$ (cm^2).
Trường hợp 2. Phương sai chưa biết, cho $t_{0.05}(40) = 1.684$, $t_{0.025}(40) = 1.684 = 2.021$,
$t_{0.05}(39) = 1.685$, $t_{0.025}(39) = 1.684 = 2.023$.

- Nếu $n \geq 120$ thì

$$P\left\{\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \leq \mu \leq \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right\} = 1 - \alpha$$

Trong đó $z_{\alpha/2}$ được tính theo phân phối chuẩn tắc N(0, 1), tức là:

$P\{Z > z_{\alpha/2}\} = \alpha/2$, với Z là BNN chuẩn tắc N(0, 1).

α	10%	5%	2%	1%	0,3%
$z_{\alpha/2}$	1,645	1,96	2,326	2,576	3

Còn S là căn bậc 2 của phương sai mẫu.

Ví dụ 5. Một trường đại học tiến hành 1 nghiên cứu xem 1 SV trung bình tiêu hết bao nhiêu tiền điện thoại trong 1 tháng. Một mẫu ngẫu nhiên gồm 59 sv được chọn và kết quả cho như 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	37	60	41	35	26	20
58	33	23	35	34	70	50	6	50	20	50
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	37	60	41	35	26	20
58	33	23	35							

Hãy xây dựng khoảng tin cậy 95% cho số tiền gọi điện thoại trung bình μ hàng tháng của 1 sv.

Giải

Từ các số liệu trên ta có: n=125, $\overline{x} = 41{,}05$, S=27.99

Do đó $\frac{S}{\sqrt{n}} = \frac{27.99}{\sqrt{125}} = 2.503$.

Vì n=125>120 nên ta có khoảng tin cậy xấp xỉ 95% cho μ là

$\overline{x} \pm 1.96(2.503) = 41.05 \pm 4.906$

Hay 36.144≤μ≤45.956

B. Khi X là BNN tùy ý:

Điều kiện: n≥120.

B1. Ước lượng khoảng của kỳ vọng khi phương sai $\sigma^2 = \sigma_0{}^2$ đã biết:

Khoảng tin cậy xấp xỉ γ cho giá trị trung bình là :

$$P\left\{\overline{x} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}} \le \mu \le \overline{x} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right\} = 1-\alpha$$

B2. Ước lượng khoảng của kỳ vọng khi phương sai σ^2 chưa biết:

Khoảng tin cậy xấp xỉ γ cho giá trị trung bình là :

$$P\left\{\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}} \le \mu \le \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\right\} = 1-\alpha$$

Ví dụ cho Trường hợp B1:

Giả sử ta có mẫu cỡ n=144 với phân phối chưa biêt nhưng biết phương sai bằng 25 cm2, ta tính được trung bình mẫu bằng 168 cm.

Xây dựng khoảng tin cậy 99% cho giá trị TB μ của X.

Giải

Ví dụ cho Trường hợp B2:

Giả sử ta có mẫu cỡ n=144 với phân phối chưa biêt và biết trung bình mẫu bằng 168 cm và $\sum_{i=1}^{n} x_i^2 = 4445000 cm^2$.

Xây dựng khoảng tin cậy 99% cho giá trị TB μ của X.

Giải

2. Ước lượng khoảng cho xác suất (Tỷ lệ):

+ Ước lượng khoảng của xác suất:

Điều kiện:

- $n \geq 30$

- $\min(n\hat{p}; n(1-\hat{p})) \geq 10.$

Hay nói cách khác, min (m, n – m) ≥0

Khoảng tin cậy xấp xỉ γ,

$$\hat{p} = f_n = \frac{m}{n}$$

$$P\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right.$$

.

Ví dụ 6. Trước cuộc bầu cử Chủ tịch Hội Phụ nữ VN, một cuộc thăm dò dư luận đã được tiến hành. Người ta chọn ngẫu nhiên 10000 người để hỏi ý kiến thì có 6000 người nói rằng họ sẽ bỏ phiếu cho ứng cử viên Nguyễn Thị Vân. Tìm khoảng tin cậy 90%, 95% và 99% cho tỷ lệ cử tri bỏ phiếu cho ứng cử viên trên.

α	10%	5%	2%	1%	0,3%
$z_{\alpha/2}$	1,645	1,96	2,326	2,576	3

Giải

Ta có n=10000; m=6000; $\hat{p} = \frac{m}{n} = 0.6$.

Ta thấy

$$n\hat{p} = 100x0.6 = 6000 > 10$$

$$n(1-\hat{p}) = 100x0.4 = 4000 > 10$$

Như vậy $\hat{p}$ sẽ có phân bố xấp xỉ chuẩn với $E(\hat{p}) = p$ với độ lệch tiêu chuẩn là

$\hat{\sigma} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = \sqrt{\frac{0.6x0.4}{10000}} = \sqrt{0.000024} = 0.0049$.

Với γ=0.95 thì $z_{\alpha/2} = 1,96$.

Vậy khoảng tin cậy 95% cho p là:

$$\hat{p} \pm 1.96x(0.0049) = 0.60 \pm 0.0096$$

Hay $0.5904 \leq p \leq 0.6096$.

3.Ước lượng khoảng cho phương sai:

+ Ước lượng khoảng của phương sai khi đã biết giá trị trung bình $\mu=\mu_0$: Khi N=n ta có công thức

$$P\left(\frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2{}_n(\alpha/2)} \leq \sigma^2 \leq \frac{\sum_{i=1}^{n}(X_i-\mu_0)^2}{\chi^2{}_n(1-\alpha/2)}\right) = 1-$$

Trong đó các giá trị $\chi_n^2(\alpha/2)$ và $\chi_n^2(1-\alpha/2)$ là các giá trị được tra từ bảng phân phối χ^2 với n bậc tự do, cụ thể là:

$$P\{\chi_n^2 > \chi_n^2(\alpha/2)\} = \alpha/2$$

$$P\{\chi_n^2 > \chi_n^2(1-\alpha/2)\} = 1-\alpha/2$$

+ Ước lượng khoảng của phương sai khi chưa biết giá trị trung bình:Khi N=n ta có công thức:

$$P\left(\frac{(n-1)s^2}{\chi^2{}_{n-1}(\alpha/2)} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2{}_{n-1}(1-\alpha/2)}\right) = 1-\alpha$$

Trong đó các giá trị $\chi_{n-1}^2(\alpha/2)$ và $\chi_{n-1}^2(1-\alpha/2)$ là các giá trị được tra từ bảng phân phối χ^2 với n-1 bậc tự do, cụ thể là:

$$P\left\{\chi^2_{n-1} > \chi^2_{n-1}(\alpha/2)\right\} = \alpha/2$$
$$P\left\{\chi^2_{n-1} > \chi^2_{n-1}(1-\alpha/2)\right\} = 1-\alpha/2$$

Ví dụ:

Giả sử ta có mẫu cỡ n=144 với phân phối chưa biết và biết trung bình mẫu bằng 168 cm và $\sum_{i=1}^{n} x_i^2 = 4445000cm^2$.

Xây dựng khoảng tin cậy 99% cho giá trị phương sai σ^2 của X.

Biết $\chi^2_{143}(0.995) = 103{,}196$, $\chi^2_{143}(0.005) = 190{,}306$.

Giải

Mệnh đề : Giả sử X_j với j=1, 2, ... là các biến ngẫu nhiên Gauss độc lập cùng phân phối, với kỳ vọng μ chưa biết và phương sai σ^2 chưa biết. Khi đó :

$(n-1)V_n^2/\sigma^2$ là biến ngẫu nhiên χ^2 với n – 1 bậc tự do.

Bài 7.3. Độ chính xác của ước lượng và cỡ mẫu cần thiết

Với độ tin cậy γ đã cho, ta thấy có mối liên quan giữa cỡ mẫu n và độ dài khoảng tin cậy. Cỡ mẫu càng lớn thì độ dài khoảng tin cậy càng hẹp, nghĩa là độ chính xác càng cao, sai số của ta càng nhỏ. Tuy nhiên, khi cỡ mẫu càng lớn thì càng đòi hỏi nhiều thời gian, tiền của, công sức.

Vậy bài toán đặt ra là: *cần chọn cỡ mẫu tối thiểu bao nhiêu để đạt được độ tin cậy và độ chính xác mong muốn.*

1. Trường hợp ước lượng cho trung bình μ:

Để ước lượng giá trị trung bình ta cần cỡ mẫu đủ lớn, cụ thể, với độ chính xác ε cho trước, ta có:

Khi phương sai đã biết

$$n_0 \geq \left(\frac{\sigma_0 \, z_{\alpha/2}}{\varepsilon}\right)^2$$

, hay khi phương sai chưa biết.

$$n_0 \geq \max\left(\left(\frac{s z_{\alpha/2}}{\varepsilon}\right)^2, 30\right)$$

Điều kiện: n≥30.

2. Trường hợp ước lượng cho tỷ lệ:

Để ước lượng xác suất với độ chính xác ε cho trước, cỡ mẫu phải đủ lớn, cụ thể là:

Cách 1: $$n_0 \geq \frac{z_{\alpha/2}^2 \hat{p}(1-\hat{p})}{\varepsilon^2}$$

Với

- $n \geq 30$

- $\min(n\hat{p}; n(1-\hat{p})) \geq 10.$

Cách 2:

$$n_0 \geq \frac{z_{\alpha/2}^2}{4\varepsilon^2}$$

- $n \geq 30$