Заняття №13

Інтервальне оцінювання невідомих параметрів розподілів

При малих об'ємах вибірки точкові оцінки невідомого параметру розподілу приводять до значних відхилень від істинного значення цього параметра і це унеможливлює їх подальше використання. Тому при малих об'ємах вибірки використовують інтервальні оцінки, які визначаються випадковими кінцями інтервалу $\hat{\theta}_1$ та $\hat{\theta}_2$.

Означення 1. Надійним інтервалом для параметра θ з рівнем надійності $1-\alpha \quad (0<\alpha<1)$ будемо називати інтервал $(\hat{\theta}_1, \hat{\theta}_2)$, який задовольняє умовам:

а) його межі $\hat{\theta}_1 = g_1(\xi_1,...,\xi_n)$, $\hat{\theta}_2 = g_2(\xi_1,...,\xi_n)$ є функціями від вибірки і не залежать від параметра θ;

б) ймовірність покриття цим інтервалом невідомого параметра θ є не меншою ніж $1-\alpha$, тобто

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) \geq 1-\alpha. \tag{1}$$

На практиці, в якості α, вибирають мале число і найчастіше одне з наступних: $0{,}1;\ 0{,}05;\ 0{,}01;\ 0{,}005;\ 0{,}001$.

Вочевидь, є слушним прагнути до того, щоб інтервал, який покриває істинне значення невідомого параметра, був як найменшої довжини, і тому умова

$$\hat{\theta}_2 - \hat{\theta}_1 \to \min$$

завжди буде братися до уваги при побудові надійного інтервалу.

З центральної граничної теореми маємо такий результат

$$P\left(\sqrt{I\left(\hat{\theta}_n\right)}\left|\hat{\theta}_n - \theta\right| < C_\alpha\right) \to \Phi(C_\alpha) - \Phi(-C_\alpha),\ n \to \infty, \tag{2}$$

де $\Phi\left(x\right) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ - функція розподілу стандартного нормального закону.

Запишемо співвідношення (2) через функцію Лапласа

$$P\left(\sqrt{I\left(\hat{\theta}_n\right)}\left|\hat{\theta}_n - \theta\right| < C_\alpha\right) \to 2\Phi_1(C_\alpha),\ n \to \infty, \tag{3}$$

де $\Phi_1\left(x\right) = \frac{1}{\sqrt{2\pi}} \int\limits_{0}^{x} e^{-\frac{t^2}{2}} dt$ - функція Лапласа.

Якщо прирівняти співвідношення (1) і (3), то будемо мати

$$2\Phi_1(C_\alpha) = 1-\alpha, \Rightarrow C_\alpha = \Phi_1^{-1}\left(\frac{1-\alpha}{2}\right).$$

Таблиця 1 для критичних величин стандартного нормального розподілу C_α:

$1-\alpha$	0,8	0,9	0,95	0,99	0,999
C_α	1,28	1,64	1,96	2,57	3,29

Цієї таблиці буде достатньо для задач на побудову надійних інтервалів, але в загальному випадку, треба вміти користуватися таблицею значень функції Лапласа

Таблиця 1. Значення функції Лапласа $\Phi_1(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	00399	00798	01197	01595	01994	02392	02790	03188	03586
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490
0,7	25804	26115	26424	26730	27035	37337	27637	27935	28230	28524
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861

3,0	0,49865	3,1	49903	3,2	49931	3,3	49952	3,4	49966
3,5	49977	3,6	49984	3,7	49989	3,8	49993	3,9	49995
4,0	499968								
4,5	499997								
5.0	49999997								

Таблиця 2. Значення $\chi^2_{k,\alpha}$ в...
= α і числа ступенів волі k.

Щільність розподілу $\chi^2(k)$

$p_{\chi^2(k)}(x) = \frac{x^{k/2-1} e^{-x/2}}{\Gamma(k/2)\, 2^{k/2}}, x > 0.$

α / k	0,99	0,95
1	0,00016	0,0039
2	0,020	0,103
3	0,115	0,352
4	0,30	0,71
5	0,55	1,14
6	1,187	1,63
7	1,24	2,17
8	1,65	2,73
9	2,09	3,32
10	2,56	3,94
11	3,1	4,6
12	3,6	5,2
13	4,1	5,9
14	4,7	6,6
15	5,2	7,3
16	5,8	8,0
17	6,4	8,7
18	7,0	9,4
19	7,6	10,1
20	8,3	10,9
21	8,9	11,6
22	9,5	12,3
23	10,2	13,1
24	10,9	13,8
25	11,5	14,6
26	12,2	15,4
27	12,9	16,2
28	13,6	16,9
29	14,3	17,7
30	15,0	18,5

У цій таблиці, в порядку зростання, записані значення функції Лапласа. Причому, мається на увазі, що перед кожним числом стоїть нуль цілих і кома. У стовпчику x записані аргументи функції Лапласа, а у рядку x записаний ще один розряд після коми для значення аргументу функції Лапласа.

Приклад 1.

Нехай рівень надійності $1-\alpha=0{,}9$. Тоді $C_\alpha = \Phi_1^{-1}\left(\frac{1-\alpha}{2}\right) = \Phi_1^{-1}(0{,}45)$. Далі в таблиці шукаємо число, яке першим перевищить $0{,}45$. Таким числом буде 45053. А нас буде цікавити число, яке передує 45053. Таким числом буде 44950. Тепер шукаємо аргумент, при якому функція Лапласа дорівнює 0,44950. По стовпчику x це буде $1{,}6$, а по рядку x це буде 4. Таким чином, шуканим аргументом функції Лапласа буде $C_\alpha = 1{,}64$.

Тепер з співвідношення (3) запишемо загальний вигляд надійного інтервалу рівня $1-\alpha$ для параметра θ

$$\hat{\theta}_n - \frac{C_\alpha}{\sqrt{I\left(\hat{\theta}_n\right)}} < \theta < \hat{\theta}_n + \frac{C_\alpha}{\sqrt{I\left(\hat{\theta}_n\right)}},$$

де $\hat{\theta}_n$ - оцінка параметра θ, знайдена за методом максимальної вірогідності;

$I\left(\hat{\theta}_n\right)$ - кількість інформації по Фішеру в точці $\hat{\theta}_n$.

Приклад 2.

На телефонній станції проводились спостереження за числом невірних з'єднань за хвилину. Результати спостережень наведено у таблиці

x_i	0	1	2	3	4	5	7
m_i	8	17	16	10	6	2	1

Відомо, що число невірних з'єднань має розподіл Пуассона з параметром θ $P\left(\xi = k\right) = \frac{\theta^k}{k!} e^{-\theta}, \theta > 0, k = 0,1,\dots$. Потрібно знайти надійний інтервал рівня $1-\alpha = 0,99$ для параметра θ.

Розв'язання.

Шукаємо об'єм вибірки

$n = \sum m_i = 60$.

Шукаємо функцію вірогідності

$$L(x,\theta) = \prod_{k=1}^{n} P(\xi = x_k) = \prod_{k=1}^{n} \frac{\theta^{x_k}}{x_k!} e^{-\theta} = \frac{\theta^{\sum_{k=1}^{n} x_k}}{\prod_{k=1}^{n} x_k!} e^{-n\theta}.$$

$$\ln L(x,\theta) = \sum_{k=1}^{n} x_k \ln\theta - n\theta - \sum_{k=1}^{n} \ln x_k!$$

$$\frac{\partial}{\partial\theta} \ln L(x,\theta) = \frac{\sum_{k=1}^{n} x_k - n\theta}{\theta} = 0.$$

Звідси маємо $\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{s} x_i m_i}{n} = \frac{120}{60} = 2$.

Далі шукаємо кількість інформації по Фішеру

$$I(\theta) = -M \frac{\partial^2}{\partial\theta^2} \ln L(\xi,\theta) = M \frac{\sum_{k=1}^{n} \xi_k}{\theta^2} = \frac{n\theta}{\theta^2} = \frac{n}{\theta} = \frac{60}{2} = 30.$$

З таблиці для критичних величин беремо $C_\alpha = 2,57$ і будуємо надійний інтервал

$$2-\frac{2,57}{\sqrt{30}}<\theta<2+\frac{2,57}{\sqrt{30}} \text{ або } 1,53<\theta<2,57.$$

Інтервальне оцінювання параметрів нормального розподілу $N(a,\sigma^2)$

При побудові надійних інтервалів для параметрів нормального розподілу, виділяють чотири випадки.

1) Надійний інтервал для математичного сподівання a, коли дисперсія σ^2 - відома

$$\bar{x}-C_\gamma\frac{\sigma}{\sqrt{n}}<a<\bar{x}+C_\gamma\frac{\sigma}{\sqrt{n}},$$

де $\bar{x}=\frac{\sum_{i=1}^{n}x_i}{n}$ - вибіркове середнє,

C_γ - критична величина стандартного нормального розподілу, яка залежить від рівня надійності та береться з наступної таблиці

Таблиця 1

$1-\gamma$	0,8	0,9	0,95	0,99	0,999
C_γ	1,28	1,64	1,96	2,57	3,29

2) Надійний інтервал для дисперсії σ^2, коли математичне сподівання a - відоме

$$\frac{nS^2}{\chi_2^2}<\sigma^2<\frac{nS^2}{\chi_1^2},$$

де $S^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-a)^2$ - вибіркова дисперсія, покращена відомим математичним сподіванням a, χ_1^2 та χ_2^2 - критичні величини χ^2- розподілу з n ступенями свободи. Ці величини беруться з таблиці критичних величин для χ^2- розподілу, в яку потрібно зайти з такими параметрами

$$P(\chi^2(n)>\chi_1^2)=\frac{2-\gamma}{2},\ P(\chi^2(n)>\chi_2^2)=\frac{\gamma}{2}.$$

Таблиця 2. Значення $\chi_{k,\alpha}^2$ в залежності від ймовірності $P(\chi^2(n)>\chi_{k,\alpha}^2)=\alpha$ і числа ступенів свободи k.

α / k	0,99	0,95	0,90	0,10	0,05	0,01
1	0,00016	0,0039	0,016	2,7	3,8	6,6
2	0,020	0,103	0,211	4,6	6,0	9,2
3	0,115	0,352	0,584	6,3	7,8	11,3
4	0,30	0,71	1,06	7,8	9,5	13,3
5	0,55	1,14	1,61	9,2	11,1	15,1
6	1,187	1,63	2,20	10,6	12,6	15,8
7	1,24	2,17	2,83	12,0	14,1	18,5
8	1,65	2,73	3,49	13,4	15,5	20,1
9	2,09	3,32	4,17	14,7	16,9	21,7
10	2,56	3,94	4,86	16,0	18,3	23,2
11	3,1	4,6	5,6	17,3	19,7	24,7
12	3,6	5,2	6,3	18,5	21,0	26,2
13	4,1	5,9	7,0	19,8	22,4	27,7
14	4,7	6,6	7,8	21,1	23,7	29,1
15	5,2	7,3	8,5	22,3	25,0	30,6
16	5,8	8,0	9,3	23,5	26,3	32,0
17	6,4	8,7	10,1	24,8	27,6	33,4
18	7,0	9,4	10,9	26,0	28,9	34,8
19	7,6	10,1	11,7	27,2	30,1	36,2
20	8,3	10,9	12,4	28,4	31,4	37,6
21	8,9	11,6	13,2	29,6	32,7	38,9
22	9,5	12,3	14,0	30,8	33,9	40,3
23	10,2	13,1	14,8	32,0	35,2	41,6
24	10,9	13,8	15,7	33,2	36,4	43,0
25	11,5	14,6	16,5	34,4	37,7	44,3
26	12,2	15,4	17,3	35,6	38,9	45,6
27	12,9	16,2	18,1	36,7	40,1	47,0
28	13,6	16,9	18,9	37,9	41,3	48,3
29	14,3	17,7	19,8	39,1	42,6	49,6
30	15,0	18,5	20,6	40,3	43,8	50,9

Приклад 3.
Задано вибірку з 10 елементів з відомим математичним сподіванням, для якої підраховано вибіркову дисперсію S^2=3. Побудувати надійний інтервал для дисперсії σ^2, якщо рівень надійності $1-\gamma=0,9$.

Розв'язання.

Щоб знайти χ_1^2 з таблиці 2, нам потрібно визначити таку ймовірність $\frac{2-\gamma}{2}$. З умови прикладу відомо, що $\gamma=0,1$. Тоді $\frac{2-\gamma}{2}=\frac{2-0,1}{2}=0,95$. Кількість ступенів свободи k дорівнює об'єму вибірки $k=n=10$. Тепер, щоб зайти в таблицю ми беремо $k=10$ та $\alpha=0,95$. Перетин відповідних рядка і стовпця дають $\chi_1^2=3,94$.

Аналогічно шукаємо χ_2^2. Тепер нам потрібно визначити ймовірність $\frac{\gamma}{2}$, виходячи з того, що рівень надійності $1-\gamma=0,9$. Тоді $\frac{\gamma}{2}=\frac{0,1}{2}=0,05$. В таблицю ми заходимо з $k=10$ та $\alpha=0,05$. Отже, $\chi_2^2=18,3$.

Можемо будувати надійний інтервал

$$\frac{10\cdot 3}{18,3}<\sigma^2<\frac{10\cdot 3}{3,94} \text{ або } 1,64<\sigma^2<7,61.$$

3) Надійний інтервал для математичного сподівання a, коли дисперсія σ^2 - невідома

$$\overline{x}-t_\gamma\frac{\hat{S}}{\sqrt{n}}<a<\overline{x}+t_\gamma\frac{\hat{S}}{\sqrt{n}},$$

де $\overline{x}=\frac{\sum_{i=1}^{n}x_i}{n}$ - вибіркове середнє,

$$\hat{S}=\sqrt{\hat{S}^2}=\sqrt{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}},$$

t_γ - критична величина розподілу Стьюдента з $n-1$ ступенем свободи, яка береться з таблиці 3, в яку входимо **з кількістю ступенів свободи** $n-1$.

4) Надійний інтервал для дисперсії σ^2, коли математичне сподівання a - невідоме

$$\frac{(n-1)\hat{S}^2}{\chi_2^2}<\sigma^2<\frac{(n-1)\hat{S}^2}{\chi_1^2},$$

де $\hat{S}^2=\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}$, χ_1^2 та χ_2^2 - критичні величини χ^2- розподілу з $n-1$ ступенем свободи. Ці величини беруться з таблиці 2 критичних величин для χ^2- розподілу, в яку потрібно зайти з такими параметрами

$$P(\chi^2(n-1)>\chi_1^2)=\frac{2-\gamma}{2},\ P(\chi^2(n-1)>\chi_2^2)=\frac{\gamma}{2}.$$

Потрібно звернути увагу, що в цьому випадку, на відміну від випадку з відомим математичним сподіванням, кількість ступенів свободи береться $n-1$.

Таблиця 3. Значення t_α для розподілу Стьюдента в залежності від ймовірності $P(|t_k|<t_\alpha)=1-\alpha$ і числа ступенів свободи k.

$1-\alpha$ / k	0,90	0,95	0,99
1	6,31	12,71	63,7
2	2,92	4,30	9,92
3	2,35	3,18	5,84
4	2,13	2,77	4,60
5	2,02	2,57	4,03
6	1,943	2,45	3,71
7	1,895	2,36	3,50
8	1,860	2,31	3,36
9	1,833	2,26	3,25
10	1,812	2,23	3,17
11	1,796	2,20	3,11
12	1,782	2,18	3,06
13	1,771	2,16	3,01
14	1,761	2,14	2,98
15	1,753	2,13	2,95
16	1,746	2,12	2,92
17	1,740	2,11	2,90
18	1,734	2,10	2,88
19	1,729	2,09	2,86
20	1,725	2,09	2,84
25	1,708	2,06	2,79
30	1,697	2,04	2,46
80	1,659	1,991	2,640
100	1,651	1,984	2,627
∞	1,645	1,960	2,576

Приклад 4.

Побудувати надійний інтервал для параметрів a і σ^2 нормального розподілу по вибірці: 0,2; 0,5; 1,0; 1,5; 0,8; 1,0; 2,0. Рівень надійності складає 0,9.

Розв'язання.

Об'єм вибірки $n = 7$. Шукаємо вибіркове середнє

$$\overline{x} = \frac{0,2+0,5+1+1,5+0,8+1+2}{7} = \frac{7}{7} = 1.$$

Шукаємо вибіркову дисперсію

$$\hat{S}^2 = \frac{1}{6}\Big((0,2-1)^2+(0,5-1)^2+(0)^2+(1,5-1)^2+(0,8-1)^2+(0)^2+(1)^2\Big) = 0,36.$$

З таблиці 3 шукаємо t_γ. Беремо кількість ступенів свободи $k = n-1 = 6$ та беремо рівень надійності $0,9$. З таблиці 3 маємо, що t_γ=1,943. Далі будуємо інтервал для a

$$1-1,943\frac{\sqrt{0,36}}{\sqrt{7}} < a < 1+1,943\frac{\sqrt{0,36}}{\sqrt{7}} \text{ або } 0,55 < a < 1,45.$$

Щоб побудувати інтервал для дисперсії, виконуємо дії аналогічні описаним у прикладі 1. Але враховуємо, що кількість ступенів свободи буде $k = n-1 = 6$. Ймовірності, з якими ми заходимо у таблицю 2, будуть $\frac{2-\gamma}{2} = \frac{2-0,1}{2} = 0,95$, $\frac{\gamma}{2} = \frac{0,1}{2} = 0,05$. Тоді відповідні критичні величини дорівнюють $\chi_1^2 = 1,63$, $\chi_2^2 = 12,6$. Надійний інтервал для дисперсії буде мати вигляд

$$\frac{6\cdot 0,36}{12,6}<\sigma^2<\frac{6\cdot 0,36}{1,63}\text{ або } 0,17<\sigma^2<1,33.$$

Розподіл χ^2 (Розподіл Пірсона).

Нехай задано n незалежних однаково розподілених випадкових величин $\xi_1,\xi_2,...,\xi_n$, таких, що кожна $\xi_i,\ i=\overline{1,n}$ розподілена за стандартним нормальним законом. Тоді $\chi^2(n)=\sum_{i=1}^{n}\xi_i^2$ буде розподілена за χ^2-розподілом з n ступенями свободи.

Розподіл Стьюдента (t-розподіл).

Нехай задано $n+1$ незалежних однаково розподілених випадкових величин $\xi_0,\xi_1,\xi_2,...,\xi_n$, таких, що кожна $\xi_i,\ i=\overline{0,n}$ розподілена за стандартним нормальним законом. Тоді $t(n)=\dfrac{\xi_0}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\xi_i^2}}=\dfrac{\xi_0}{\sqrt{\frac{1}{n}\chi^2(n)}}$ буде мати розподіл Стьюдента з n ступенями свободи.

ЗАДАЧІ

1) Побудувати надійний інтервал для параметрів a і σ^2 нормального розподілу по вибірці: 0,6; 2,4; 2,1; 1,4; 1,2; 4,8; 0,9; 1,1; 3,5; 3,0. Рівень надійності складає 0,9.