Розділ 9

СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ РОЗПОДІЛУ

9.1. Точкові оцінки

Статистичною оцінкою θ^* невідомого параметра θ теоретичного розподілу називають функцію $f(X_1, X_2, \ldots, X_n)$, де $X_1,\ X_2,\ \ldots,\ X_n$ – спостережувальні випадкові величини.

Точковою оцінкою називають статистичну оцінку, яка визначається одним числом $\theta^* = f(x_1, x_2, \ldots, x_n)$, де $x_1, x_2, \ldots, x_n$ – результати n спостережень над кількісною ознакою випадкової величини *X*.

Наприклад, вибіркова середня $\bar{x}$, вибіркова дисперсія $D_в$ та вибіркове середньоквадратичне $\sigma_в$ – точкові оцінки відповідних числових характеристик генеральної сукупності.

Точкові оцінки параметрів розподілу є випадковими величинами, їх можна вважати первинними результатами обробки вибірки тому, що невідомо, з якою точністю кожна з них оцінює відповідну числову характеристику генеральної сукупності.

Якщо об'єм вибірки доволі великий, то точкові оцінки задовольняють практичні потреби точності.

Щоб статистичні оцінки давали найкращі наближення, вони повинні задовольняти певним вимогам. Розглянемо ці вимоги.

Нехай θ^* – статистична оцінка невідомого параметра θ теоретичного розподілу. Припустимо, що за вибіркою об'єму n знайдена оцінка θ_1^*. Для інших вибірок того ж об'єму одержано оцінки $\theta_2^*,\ \theta_3^*,\ \ldots,\ \theta_m^*$. Оцінку θ^* можна розглядати як випадкову величину, а числа $\theta_1^*,\ \theta_2^*,\ \ldots,\ \theta_m^*$ – як її можливі значення.

Якщо θ_k^* $(k = 1, 2, \ldots, m)$ будуть більші від значення θ, то оцінка θ^* дає наближене значення θ з надлишком. У цьому випадку математичне сподівання випадкової величини θ^* є більшим від θ ($\mathrm{M}(\theta^*) > \theta$).

Якщо θ^* дає оцінку з недостачею, математичне сподівання випадкової величини θ^* є меншим від θ ($\mathrm{M}(\theta^*) < \theta$).

Відтак використання статистичної оцінки, математичне сподівання якої не дорівнює параметру θ, спричиняє систематичні похибки. Вимога $M(\theta^*) = \theta$ застерігає від таких похибок.

Незміщеною називають точкову оцінку θ^*, математичне сподівання якої дорівнює оцінюваному параметру θ для будь-якого обсягу вибірки:

$$M(\theta^*) = \theta.$$

Зміщеною називають точкову оцінку θ^*, математичне сподівання якої відмінне від оцінюваного параметра θ.

Якщо $M(\theta^*) = \theta + b(\theta)$, то $b(\theta)$ називають **зсувом оцінки** $\boldsymbol{\theta}^*$.

Вимога про незміщеність оцінки θ^* є недостатньою, оскільки можливі значення θ^* можуть бути дуже розсіяні від середнього значення, дисперсія $D(\theta^*)$ може бути великою. Тоді, знайдена за даними однієї вибірки оцінка θ_k^* може значно відрізнятися від середнього значення θ^*, а отже, і від параметра θ.

Якщо дисперсія $D(\theta^*)$ буде незначною, то можливість допустити помилку виключена. Тому для статистичної оцінки виникає вимога про її ефективність.

Ефективною називають таку оцінку θ^*, яка при заданому об'ємі n має найменшу можливу дисперсію.

Обґрунтованою називають статистичну оцінку, яка при $n \to \infty$ прямує за ймовірністю до оцінюваного параметра.

Послідовність оцінок θ_k^* $(k = 1, 2, \dots, m)$ параметра θ називається ***спроможною***, якщо $\theta_k^* \xrightarrow{P} \theta,\ n \to \infty$.

Послідовність оцінок θ_k^* $(k = 1, 2, \dots, m)$ параметра θ називається **сильно спроможною**, якщо $\theta_k^* \to \theta$ з ймовірністю одиниця при $n \to \infty$.

Незміщеною оцінкою генерального середнього (математичного сподівання) є вибіркове середнє

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i,$$

де n_i – частота варіанти x_i; x_i – варіанта вибірки; $n = n_1 + n_2 + \dots + n_m$ – об'єм вибірки.

Зауваження 1. Якщо варіанти x_i вибірки є дуже великими або дуже малими (близькими до нуля) числами, то для спрощення розрахунків доцільно відняти (у разі великих від'ємних чисел – додати) від кожної варіанти одне й те саме число C (як C можна вибрати будь – яке число, розміщене приблизно посередині варіаційного ряду), потім поділити (у разі близьких до нуля чисел – помножити) на одне й те саме число b (як b можна вибрати найбільше спільне кратне), тобто перейти до умовних варіант

$$u_i = \frac{x_i \pm C}{b} \qquad (u_i = x_i b). \tag{9.1}$$

Тоді

$$\bar{x} = \frac{b}{n}\sum_{i=1}^{m} n_i u_i \mp C = b\bar{u}_b \mp C \quad \left(\bar{x} = \frac{\sum_{i=1}^{m} n_i u_i}{bn} = \frac{\bar{u}_b}{b}\right). \tag{9.2}$$

Зміщеною оцінкою генеральної дисперсії є вибіркова дисперсія

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{m} n_i \left(x_i - \bar{x}\right)^2.$$

Ця оцінка зміщена, оскільки

$$M(D_{\text{в}}) = \frac{n-1}{n} D_{\Gamma}.$$

Для обчислення вибіркової дисперсії можна скористатися зручнішою формулою

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i x_i\right)^2 = \overline{x^2} - (\bar{x})^2.$$

Зауваження 2. Якщо варіанти x_i вибірки є дуже великими або дуже малими (близькими до нуля) числами, то для спрощення розрахунків доцільно відняти (у разі великих від'ємних чисел – додати) від кожної варіанти одне й те саме число С (як С можна вибрати будь-яке число, розміщене приблизно посередині варіаційного ряду), потім поділити (у разі близьких до нуля чисел – помножити) на одне й те саме число b (як b можна вибрати найбільше спільне кратне), тобто перейти до умовних варіант

$$u_i = \frac{x_i \pm C}{b} \qquad (u_i = x_i b).$$

Тоді

$$D_{\text{в}} = b^2\left(\frac{1}{n}\sum_{i=1}^{m} n_i u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i u_i\right)^2\right) = b^2\left(\overline{u^2} - (\bar{u})^2\right). \tag{9.3}$$

$$\left(D_{\text{в}} = \left(\frac{1}{b^2 n}\sum_{i=1}^{m} n_i u_i^2 - \left(\frac{1}{bn}\sum_{i=1}^{m} n_i u_i\right)^2\right) = \frac{\left(\overline{u^2} - (\bar{u})^2\right)}{b^2}\right). \tag{9.4}$$

Незміщеною оцінкою генеральної дисперсії є виправлена вибіркова дисперсія

$$S^2 = \frac{n}{n-1} D_{\text{в}}.$$

Поправку $\frac{n}{n-1}$ називають ***поправкою Бесселя***.

Приклад 9.1. З генеральної сукупності отримано деяку вибірку об'єму $n = 100$:

x_i	0,002	0,004	0,006	0,008	0,01	0,012	0,014
n_i	7	29	35	12	9	5	3

Знайти зміщену оцінку генеральної дисперсії.

Розв'язання. Зміщеною оцінкою генеральної дисперсії D_{Γ} є вибіркова дисперсія $D_{\text{в}}$. Оскільки варіанти вибірки є малими, близькими до нуля числами, перейдемо до умовних варіант

$$u_i = x_i \cdot 1000$$

(як b вибрано число 1000, оскільки в такому разі ми отримаємо цілі числа):

$$u_1 = x_1 \cdot 1000 = 0{,}002 \cdot 1000 = 2;$$
$$u_2 = x_2 \cdot 1000 = 0{,}004 \cdot 1000 = 4;$$
$$u_3 = x_3 \cdot 1000 = 0{,}006 \cdot 1000 = 6;$$
$$u_4 = x_4 \cdot 1000 = 0{,}008 \cdot 1000 = 8;$$
$$u_5 = x_5 \cdot 1000 = 0{,}01 \cdot 1000 = 10;$$
$$u_6 = x_6 \cdot 1000 = 0{,}012 \cdot 1000 = 12;$$
$$u_7 = x_7 \cdot 1000 = 0{,}014 \cdot 1000 = 14.$$

Тепер можна знайти вибіркову дисперсію:

$$D_{\text{в}} = \left(\frac{1}{b^2 n} \sum_{i=1}^{7} n_i u_i^2 - \left(\frac{1}{bn} \sum_{i=1}^{7} n_i u_i\right)^2\right) =$$

$$= \frac{7 \cdot 2^2 + 29 \cdot 4^2 + 35 \cdot 6^2 + 12 \cdot 8^2 + 9 \cdot 10^2 + 5 \cdot 12^2 + 3 \cdot 14^2}{1000^2 \cdot 100} -$$

$$-\left[\frac{7\cdot 2+29\cdot 4+35\cdot 6+12\cdot 8+9\cdot 10+5\cdot 12+3\cdot 14}{1000\cdot 100}\right]^2=$$

$$=\frac{28+464+1260+768+900+720+588}{100000000}-$$

$$-\left[\frac{14+116+210+96+90+60+42}{100000}\right]^2=$$

$$=4728\cdot 10^{-8}-394384\cdot 10^{-10}=78416\cdot 10^{-10}=0{,}0000078416.$$

Приклад 9.2. У результаті статистичних досліджень випадкової величини X отримано таку вибірку: 47, 45, 46, 46, 46, 45, 47, 44, 46, 45, 45, 46, 46, 44, 46, 48, 46, 46, 45, 46, 44, 46, 45, 47, 46, 46, 47, 46, 46, 48, 44, 46, 45, 46, 45, 46, 44, 47, 46, 46, 45, 47, 48, 44, 46, 46, 45, 46, 47, 45.

Знайти незміщені оцінки генерального середнього та генеральної дисперсії.

Розв'язання. Об'єм вибірки $n=50$. Побудуємо статистичний розподіл вибірки:

x_i	44	45	46	47	48
n_i	6	11	23	7	3

Перевірка: $n = 6 + 11 + 23 + 7 + 3 = 50.$

Незміщеною оцінкою генерального середнього є вибіркове середнє.

Оскільки варіанти вибірки є великими числами, перейдемо до умовних варіант

$$u_i = x_i - 46.$$

Як С вибрано число 46:

$$u_1 = x_1 - 46 = 44 - 46 = -2; \qquad u_2 = x_2 - 46 = 45 - 46 = -1;$$

$$u_3 = x_3 - 46 = 46 - 46 = 0; \qquad u_4 = x_4 - 46 = 47 - 46 = 1;$$

$$u_5 = x_5 - 46 = 48 - 46 = 5.$$

Тепер можна знайти вибіркове середнє:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{m} n_i u_i + C = \frac{-2\cdot6+(-1)\cdot11+0\cdot23+1\cdot7+2\cdot3}{50} + 46 =$$
$$= \frac{-10}{50} + 46 = 45{,}8.$$

Щоб знайти незміщену оцінку генеральної дисперсії (виправлену вибіркову дисперсію), визначимо вибіркову дисперсію й помножимо її на поправку Бесселя:

$$D_{\text{в}} = \frac{1}{n}\sum_{i=1}^{m} n_i x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i x_i\right)^2 =$$

$$= \frac{(-2)^2\cdot6 + (-1)^2\cdot11 + 0^2\cdot23 + 1^2\cdot7 + 2^2\cdot3}{50} -$$

$$-\left[\frac{-2\cdot6 + (-1)\cdot11 + 0\cdot23 + 1\cdot7 + 2\cdot3}{50}\right]^2 =$$

$$= \frac{54}{50} - \left[\frac{-10}{50}\right]^2 = 1{,}04.$$

$$S^2 = \frac{n}{n-1}D_{\text{в}} = \frac{50}{49}\cdot1{,}04 \approx 1{,}06.$$

Отже, незміщеними оцінками генерального середнього та генеральної дисперсії є

$$\bar{x}_{\Gamma}^{*} = 45{,}8; \qquad \bar{D}_{\Gamma}^{*} \approx 1{,}06.$$

9.2. Методи визначення точкових статистичних оцінок

9.2.1. Метод моментів

Методом моментів знаходження точкових оцінок називають метод, за яким для обчислення невідомих параметрів заданого розподілу прирівнюють теоретичні та емпіричні моменти.

Якщо розподіл визначається одним параметром, то для знаходження оцінки прирівнюють математичне сподівання до вибіркового середнього

$$M(X) = \bar{x},$$

а потім із цього рівняння визначають шукану точкову оцінку невідомого параметра.

Якщо розподіл визначається двома параметрами, то їх точкові оцінки знаходять із системи рівнянь:

$$\begin{cases} M(X) = \bar{x}; \\ D(X) = D_{\text{в}}. \end{cases} \quad (9.5)$$

Лівими частинами цих рівнянь є математичне сподівання та дисперсія, які дорівнюють відповідно вибірковому середнього та вибірковій дисперсії.

Приклад 9.3. Випадкова величина *X* (зріст дорослої людини) розподілена за нормальним розподілом з параметрами *a, σ.* У результаті статистичних досліджень отримано такий статистичний розподіл зросту дорослих людей для $n = 1000$ осіб:

Зріст, см	(145; 155]	(155; 165]	(165; 175]	(175; 185]	(185; 195]	(195; 205]	(205; 215]
Кількість осіб	24	112	263	322	202	66	11

Знайти методом моментів точкову оцінку невідомих параметрів *a, σ* нормального розподілу.

Розв'язання. Перетворимо інтервальний статистичний розподіл на точковий, вибравши як варіанти середини частинних інтервалів:

Зріст, см	150	160	170	180	190	200	210
Кількість осіб	24	112	263	322	202	66	11

Оскільки нормальний закон розподілу залежить від двох параметрів, треба знайти вибіркове середнє та вибіркову дисперсію, а потім прирівняти їх відповідно до математичного сподівання та дисперсії.

Перейдемо до умовних варіант

$$u_i = \frac{x_i - 180}{10}$$

(x_i – зріст людини):

$$u_1 = \frac{150-180}{10} = -3; \quad u_2 = \frac{160-180}{10} = -2; \quad u_3 = \frac{170-180}{10} = -1;$$

$$u_4 = \frac{180-180}{10} = 0; \quad u_5 = \frac{190-180}{10} 1; \quad u_6 = \frac{200-180}{10} = 2;$$

$$u_7 = \frac{210-180}{10} = 3.$$

Знайдемо вибіркове середнє та вибіркову дисперсію:

$$\bar{x} = \frac{10}{n}\sum_{i=1}^{m} n_i u_i + 180 =$$

$$= \frac{10\cdot(-3\cdot 24 + (-2)\cdot 112 + (-1)\cdot 263 + 0\cdot 322 + 1\cdot 202 + 2\cdot 66 + 3\cdot 11)}{1000} + 180 =$$

$$= 0{,}01 \cdot (-72 - 224 - 263 + 0 + 202 + 132 + 33) + 180 =$$

$$= 0{,}01 \cdot (-192) + 180 = 178{,}08;$$

$$D_{\text{в}} = b^2\left(\frac{1}{n}\sum_{i=1}^{m} n_i u_i^2 - \left(\frac{1}{n}\sum_{i=1}^{m} n_i u_i\right)^2\right) =$$

$$= 10^2 \cdot \left(\frac{(-3)^2\cdot 24 + (-2)^2\cdot 112 + (-1)^2\cdot 263 + 0^2\cdot 322 + 1^2\cdot 202 + 2^2\cdot 66 + 3^2\cdot 11}{1000}\right) -$$

$$- \left[\frac{(-3)\cdot 24 + (-2)\cdot 112 + (-1)\cdot 263 + 0\cdot 322 + 1\cdot 202 + 2\cdot 66 + 3\cdot 11}{1000}\right]^2 =$$

$$= 100 \cdot \left(\frac{1492}{1000} - \left[\frac{-192}{1000}\right]^2\right) = 100 * 1{,}455136 = 145{,}5136.$$

Оскільки параметр a нормального закону розподілу є математичним сподіванням, а параметр σ – середнім квадратичним відхиленням, то оцінками цих параметрів є

$$a^* = \bar{x} = 178{,}08; \qquad \sigma^* = \sqrt{D_{\text{в}}} = \sqrt{145{,}5136} \approx 12{,}0629.$$

9.2.2. Метод найменших квадратів

Згідно з цим методом статистичні оцінки визначаються з умови мінімізації суми квадратів відхилень варіант вибірки від статистичної оцінки θ^*.

Отож, використовуючи метод найменших квадратів, можна, наприклад, визначити статистичну оцінку для $\overline{X_\Gamma} = M(X)$. Для цього скористаємося функцією $u = \sum_{i=1}^{n}\left(x_i - \theta^*\right)^2 n_i$. Використовуючи умову екстремуму, отримаємо:

$$\frac{\partial u}{\partial \theta^*} = -2\sum_{i=1}^{n}\left(x_i - \theta^*\right)n_i = 0 \rightarrow$$

$$\rightarrow \sum_{i=1}^{n} x_i n_i - \sum_{i=1}^{n} \theta^* n_i = 0 \rightarrow$$

$$\theta^* = \frac{\sum_{i=1}^{n} x_i n_i}{n} = \overline{x_\text{B}}.$$

Звідси для $\theta = \overline{X_\Gamma}$ точковою статистичною оцінкою буде $\theta^* = \bar{x}$ – вибіркова середня.

9.2.3. Метод максимальної правдоподібності

Метод максимальної правдоподібності займає головне місце в теорії статистичного оцінювання параметрів θ. На нього свого часу зауважував К. Гаусс, а розробив його Р. Фішер. Цей метод розглянемо детальніше.

Метод максимальної правдоподібності полягає у знаходженні максимуму функції одного або кількох оцінюваних параметрів.

Припустимо, що X – дискретна випадкова величина з відомим законом розподілу, який визначається невідомим параметром θ. За даними вибірки $x_1, x_2, \dots, x_n$, отриманої в результаті спостережень над випадковою величиною X, необхідно знайти точкову оцінку $\theta^* = \theta^*(x_1, x_2, \dots, x_n)$ параметра θ.

Функцією правдоподібності L дискретної випадкової величини називають функцію аргумента θ

$$L(x_1, x_2, \dots, x_n; \theta) = p(x_1; \theta) \cdot p(x_2; \theta) \cdot \dots \cdot p(x_n; \theta), \qquad (9.6)$$

де $p(x_i; \theta)$ – ймовірність того, що в результаті випробування випадкова величина X набуде значення x_i.

Оцінкою найбільшої або максимальної правдоподібності параметра θ називають таке його значення θ^*, за якого функція правдоподібності досягає свого максимуму.

Логарифмічною функцією правдоподібності називають функцію $\ln L$.

Оскільки функції L і $\ln L$ досягають свого максимуму при одному і тому самому значенні θ, здебільшого зручніше знаходити максимум функції $\ln L$, а не L.

Якщо випадкова величина *X* неперервна, то відомою вважається щільність розподілу ймовірностей $f(x)$, а невідомим – параметр, від якого залежить ця щільність.

Функцією правдоподібності L неперервної випадкової величини називають функцію аргумента θ

$$L(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta). \qquad (9.7)$$

Оцінку найбільшої правдоподібності невідомого параметра розподілу неперервної випадкової величини шукають так само, як і у випадку дискретної випадкової величини, а саме:

1) визначають похідну $\frac{\partial \ln L}{\partial \theta}$ (або $\frac{\partial L}{\partial \theta}$);

2) знаходять корені θ_i^* рівняння $\frac{\partial \ln L}{\partial \theta} = 0$ (або $\frac{\partial L}{\partial \theta} = 0$). Ці рівняння називають ***рівняннями правдоподібності***;

3) визначають другу похідну $\frac{\partial^2 \ln L}{\partial \theta^2}$ (або $\frac{\partial^2 L}{\partial \theta^2}$). Корінь θ_i^* рівняння правдоподібності, для якого друга похідна від'ємна, беруть за оцінку θ^* найбільшої правдоподібності параметра θ.

Якщо параметр θ двовимірний, тобто $\theta = (\theta_1, \theta_2)$, то для знаходження максимуму функції правдоподібності складають і розв'язують систему рівнянь:

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_1} = 0; \\ \frac{\partial \ln L}{\partial \theta_2} = 0. \end{cases} \quad \text{або} \quad \begin{cases} \frac{\partial L}{\partial \theta_1} = 0; \\ \frac{\partial L}{\partial \theta_2} = 0. \end{cases} \qquad (9.8)$$

Наприклад, коли ознака генеральної сукупності *X* має нормальний закон розподілу, то функція максимальної правдоподібності набере такого вигляду:

$$L(x_1, x_2, \dots, x_n; \theta_1^*, \theta_2^*) = \frac{1}{(2\pi\theta_2)^{\frac{n}{2}}} \cdot e^{-\frac{\sum_{i=1}^{n}(x_i - \theta_1^*)^2}{2\theta_2^*}}. \qquad (9.9)$$

Водночас за статистичні оцінки θ_1^*, θ_2^* вибирають ті їх значення, за яких задана вибірка буде найімовірнішою, тобто функція $L(x_1, x_2, \dots, x_n; \theta_1^*, \theta_2^*)$ досягає максимуму.

На практиці зручно від функції (9.9) перейти до її логарифма, а саме:

$$\ln L(x_1, x_2, \dots, x_n; \theta_1^*, \theta_2^*) = -\frac{n}{2}(\ln \pi + \ln \theta_2^*) - \frac{\sum_{i=1}^{n}(x_i - \theta_1^*)^2}{2\theta_2^*}. \qquad (9.10)$$

Згідно з необхідною умовою екстремуму для цієї функції дістанемо:

$$\begin{cases} \frac{\partial \ln L}{\partial \theta_1^*} = -\frac{\sum_{i=1}^{n}(x_i - \theta_1^*)}{\theta_2^*} = 0; \\ \frac{\partial \ln L}{\partial \theta_2^*} = -\frac{n}{2\theta_2^*} + \frac{\sum_{i=1}^{n}(x_i - \theta_1^*)^2}{2(\theta_2^*)^2} = 0. \end{cases} \quad (9.11)$$

З першого рівняння системи дістанемо:

$$\theta_1^* = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

З другого рівняння системи маємо:

$$\theta_2^* = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = D_{\text{в}}.$$

Отже, для $\overline{X_{\Gamma}} = M(X)$ точковою статистичною оцінкою є $\bar{x}$ для $D_{\Gamma} - D_{\text{в}}$.

Приклад 9.4. Випадкова величина X (кількість битого скляного посуду в одній упаковці) розподілена за законом Пуассона з невідомим параметром λ. У результаті статистичних досліджень отримано такий емпіричний розподіл кількості битого скляного посуду в $n = 1000$ упаковках:

x_i	0	1	2	3	4	5	6
n_i	554	324	98	19	3	1	1

Знайти методом максимальної правдоподібності точкову оцінку невідомого параметра λ розподілу Пуассона.

Розв'язання. Оскільки випадкова величина X розподілена за законом Пуассона, то функція правдоподібності має вигляд:

$$L = \left(\frac{\lambda^0 e^{-\lambda}}{0!}\right)^{554} \cdot \left(\frac{\lambda^1 e^{-\lambda}}{1!}\right)^{324} \cdot \left(\frac{\lambda^2 e^{-\lambda}}{2!}\right)^{98} \cdot \left(\frac{\lambda^3 e^{-\lambda}}{3!}\right)^{19} \cdot \left(\frac{\lambda^4 e^{-\lambda}}{4!}\right)^{3} \cdot \left(\frac{\lambda^5 e^{-\lambda}}{5!}\right)^{1} \cdot \left(\frac{\lambda^6 e^{-\lambda}}{6!}\right)^{1}.$$

Запишемо логарифмічну функцію правдоподібності:

$$\ln L = 600 \ln \lambda - 1000\lambda - \ln(2^{133} \cdot 3^{25} \cdot 5^2).$$

Визначимо похідну логарифмічної функції розподілу за λ:

$$\frac{\partial \ln L}{\partial \lambda} = \frac{600}{\lambda} - 1000.$$

Прирівнявши її до нуля, знайдемо єдиний корінь рівняння правдоподібності:

$$\lambda^* = 0{,}6.$$

Оскільки друга похідна логарифмічної функції правдоподібності

$$\frac{\partial^2 \ln L}{\partial \lambda^2} = -\frac{600}{\lambda^2} < 0$$

завжди від'ємна, точковою оцінкою максимальної правдоподібності параметра λ розподілу Пуассона буде

$$\lambda^* = 0{,}6.$$

9.3. Інтервальні оцінки

Якщо об'єм вибірки малий, то точкові оцінки можуть давати значні похибки, тому питання точності оцінок у цьому випадку дуже важливе і використовують інтервальні оцінки.

Інтервальною називають статистичну оцінку, що визначається двома числами, кінцями інтервалів.

Нехай знайдена за даними вибірки статистична оцінка θ^* буде оцінкою невідомого параметра θ.

Різниця між статистичною оцінкою θ^* та її оцінювальним параметром θ, взята за абсолютним значенням, називається *точністю оцінки*, а саме:

$$\left|\theta^* - \theta\right| < \delta, \tag{9.12}$$

де δ є точністю оцінки.

Оскільки θ^* є випадковою величиною, то і δ буде випадковою, тому нерівність (9.12) справджуватиметься з певною ймовірністю.

Імовірність, з якою береться нерівність (9.12), тобто

$$P\left(\left|\theta^* - \theta\right| < \delta\right) = \gamma \tag{9.13}$$

називають *надійністю*.

Рівність (9.13) можна записати так:

$$P\left(\theta^{*} - \delta < \theta < \theta^{*} + \delta\right) = \gamma \tag{9.14}$$

Найчастіше число γ задається наперед і, залежно від обставин, воно дорівнює 0,95 або 0,99 або 0,999.

Інтервал $\left[\theta^{*} - \delta;\ \theta^{*} + \delta\right]$, що покриває оцінюваний параметр θ генеральної сукупності зі заданою надійністю γ, називають *довірчим*.

Зауваження. Кінці довірчого інтервалу є випадковими величинами.

9.3.1. Побудова довірчого інтервалу для $\overline{X_\Gamma}$ при відомому значенні σ_Γ із заданою надійністю γ

Нехай ознака X генеральної сукупності має нормальний закон розподілу. Побудуємо довірчий інтервал для $\overline{X_\Gamma}$, знаючи числове значення середнього квадратичного відхилення генеральної сукупності σ_Γ; із заданою надійністю γ. Оскільки $\bar{x}$ як точкова незміщена статистична оцінка для $\overline{X_\Gamma} = M(X)$ має нормальний закон розподілу з числовими характеристиками $M(\bar{x}) = \alpha$, $\sigma(\bar{x}) = \frac{\sigma_\Gamma}{\sqrt{n}}$, то, скориставшись (9.14), дістанемо

$$P\left(\left|\bar{x} - \alpha\right| < \delta\right) = \gamma. \tag{9.15}$$

Випадкова величина $\bar{x} - \alpha$ має нормальний закон розподілу з числовими характеристиками

$$M\left(\bar{x} - \alpha\right) = M(\bar{x}) - \alpha = \alpha - \alpha = 0,$$

$$D\left(\bar{x} - \alpha\right) = D(\bar{x}) = \frac{D_\Gamma}{n},$$

$$\sigma(\bar{x}) = \frac{\sigma_\Gamma}{\sqrt{n}}.$$

Тому $\frac{\bar{x} - \alpha}{\frac{\sigma_\Gamma}{\sqrt{n}}}$ матиме нормований нормальний закон розподілу $N(0;\ 1)$.

Звідси рівність (9.15) можна записати, назначивши $\frac{\delta}{\frac{\sigma_\Gamma}{\sqrt{n}}} = x$ так:

$$P\left(\left|\frac{\bar{x} - \alpha}{\frac{\sigma_\Gamma}{\sqrt{n}}}\right| < x\right) = \gamma \tag{9.16}$$

або

$$P\left(\bar{x}-\frac{x\cdot\sigma_Г}{\sqrt{n}}<a<\bar{x}+\frac{x\cdot\sigma_Г}{\sqrt{n}}\right)=\gamma.$$

Згідно з формулою нормованого нормального закону

$$P(|X-a|<\delta)=2\Phi(\delta)$$

для (9.16) вона набирає такого вигляду:

$$P\left(\left|\frac{\bar{x}-\alpha}{\frac{\sigma_Г}{\sqrt{n}}}\right|<x\right)=2\Phi(x)=\gamma. \qquad (9.17)$$

З рівності (9.17) знаходимо аргументи *x*, а саме:

$$2\Phi(x)=\gamma\rightarrow\Phi(x)=\frac{\gamma}{2}.$$

Аргумент *x* знаходимо за значенням функції Лапласа, яка дорівнює $\frac{1}{2}\gamma$ за таблицею.

Отже, довірчий інтервал дорівнюватиме:

$$\bar{x}-\frac{x\cdot\sigma_Г}{\sqrt{n}}<a<\bar{x}+\frac{x\cdot\sigma_Г}{\sqrt{n}}. \qquad (9.18)$$

Величина $\frac{x\cdot\sigma_Г}{\sqrt{n}}$ називається *точністю оцінки* або *похибкою вибірки*.

Приклад 9.5. Вимірявши 40 випадково відібраних після виготовлення деталей, знайшли вибіркову середню, що дорівнює 15 см. Із надійністю $\gamma=0{,}99$ побудувати довірчий інтервал для середньої величини всієї партії деталей, якщо генеральна дисперсія дорівнює 0,99 см^2.

Розв'язання. Для побудови довірчого інтервалу необхідно знати: $\bar{x},\ \sigma_Г,\ n,\ x.$

З умови задачі маємо: $\bar{x}=15$ см, $\sigma_Г=\sqrt{D_Г}=\sqrt{0{,}99}=0{,}3$ см, $n=40$. Величина *x* обчислюється з рівняння:

$$\Phi(x)=\frac{\gamma}{2}=\frac{0{,}99}{2}=0{,}495;$$

$\Phi(x)=0{,}495\ \rightarrow x=2{,}58$ (за таблицею значень функції Лапласа).

Знайдемо числові значення кінців довірчого інтервалу:

$$\bar{x}-\frac{x\cdot\sigma_\Gamma}{\sqrt{n}} = 15-\frac{2{,}58\cdot0{,}3}{\sqrt{40}} = 15-0{,}12 = 14{,}88 \text{ см;}$$

$$\bar{x}+\frac{x\cdot\sigma_\Gamma}{\sqrt{n}} = 15+\frac{2{,}58\cdot0{,}3}{\sqrt{40}} = 15+0{,}12 = 15{,}12 \text{ см.}$$

Відтак маємо:

$$14{,}88 < \overline{X_\Gamma} < 15{,}12.$$

Отже, з надійністю 0,99 (99% гарантії) оцінюваний параметр $\overline{X_\Gamma}$ перебуває всередині інтервалу $[14{,}87;\ 15{,}13]$.

Приклад 9.6. Маємо такі дані про розміри основних фондів (у млн грн) на 30-ти випадково вибраних підприємствах:

4,2; 2,4; 4,9; 6,7; 4,5; 2,7; 3,9; 2,1; 5,8; 4,0;
2,8; 7,8; 4,4; 6,6; 2,0; 6,2; 7,0; 8,1; 0,7; 6,8;
9,4; 7,6; 6,3; 8,8; 6,5; 1,4; 4,6; 2,0; 7,2; 9,1.

Побудувати інтервальний статистичний розподіл із довжиною кроку h = 2 млн грн.

З надійністю $\gamma = 0{,}999$ знайти довірчий інтервал для $\overline{X_\Gamma}$, якщо σ_Γ = 5 млн грн.

Розв'язання. Інтервальний статистичний розподіл буде таким:

h = 2 млн грн.	2 – 4	4 – 6	6 – 8	8 – 10
n_i	9	7	10	4

Для визначення $\bar{x}$ необхідно побудувати дискретний статистичний розподіл, що має такий вигляд:

x_i^*	3	5	7	9
n_i	9	7	10	4

$$n = \sum n_i = 30.$$

Тоді

$$\bar{x} = \frac{\sum n_i x_i^*}{n} = \frac{3\cdot 9+5\cdot 7+7\cdot 10+9\cdot 4}{30} = \frac{168}{30} = 5{,}6 \text{ млн грн.}$$

Для побудови довірчого інтервалу із заданою надійністю $\gamma = 0{,}999$ необхідно знайти *x*:

$$\Phi(x) = \frac{\gamma}{2} = \frac{0{,}999}{2} = 0{,}4995;$$

$\Phi(x) = 0{,}4995 \ \rightarrow x = 3{,}4$ (за таблицею значень функції Лапласа).

Обчислюємо кінці інтервалу:

$$\bar{x} - \frac{x\cdot\sigma_\Gamma}{\sqrt{n}} = 5{,}6 - \frac{3{,}4\cdot 5}{\sqrt{30}} = 5{,}6 - 3{,}1 = 2{,}5 \text{ млн грн;}$$

$$\bar{x} + \frac{x\cdot\sigma_\Gamma}{\sqrt{n}} = 5{,}6 + \frac{3{,}4\cdot 5}{\sqrt{30}} = 5{,}6 + 3{,}1 = 8{,}7 \text{ млн грн.}$$

Отже, довірчий інтервал для $\overline{X_\Gamma}$ буде $2{,}5 < \overline{X_\Gamma} < 8{,}7$.

9.3.2. Побудова довірчого інтервалу для $\overline{X}_\Gamma$ при невідомому значенні σ_Γ із заданою надійністю γ

Для малих вибірок, з якими стикаємося, досліджуючи різні ознаки в економіці, техніці чи сільському господарстві, для оцінювання $\overline{X_\Gamma} = a$ при невідомому значенні σ_Γ неможливо скористатися нормальним законом розподілу. Тому для побудови довірчого інтервалу застосовується випадкова величина

$$t = \frac{\bar{x} - a}{\frac{S}{\sqrt{n}}}, \tag{9.19}$$

що має розподіл Стьюдента з $k = n - 1$ ступенями свободи.

Тоді (9.19) набуває такого вигляду:

$$P\left(\left|\frac{\bar{x}-a}{\frac{S}{\sqrt{n}}}\right| < t_\gamma\right) = P\left(\bar{x} - \frac{t_\gamma\cdot S}{\sqrt{n}} < a < \bar{x} + \frac{t_\gamma\cdot S}{\sqrt{n}}\right) = 2\int_0^{t_\gamma} f(t)dt = \gamma,$$

оскільки $f(t)$ для розподілу Стьюдента є функцією парною.

Обчисливши за цим статистичним розподілом $\bar{x}$, S і визначивши за таблицею розподілу Стьюдента значення t_γ, будуємо довірчий інтервал

$$\bar{x} - \frac{t_\gamma \cdot S}{\sqrt{n}} < a < \bar{x} + \frac{t_\gamma \cdot S}{\sqrt{n}}. \qquad (9.20)$$

Тут $t_\gamma(\gamma; k = n - 1)$ обчислюємо за заданою надійністю γ і числом ступенів свободи $k = n - 1$ за таблицею.

Приклад 9.7. Випадково вибрана партія з двадцяти приладів була випробувана щодо терміну безвідказної роботи кожного з них t_i. Результати випробувань наведено у вигляді дискретного статистичного розподілу:

t_i	100	170	240	310	380
n_i	2	5	10	2	1

З надійністю $\gamma = 0{,}99$ побудувати довірчий інтервал для «a» (середнього часу безвідказної роботи приладу).

Розв'язання. Для побудови довірчого інтервалу необхідно знайти середнє вибіркове і виправлене середнє квадратичне відхилення.

Обчислимо $\bar{x}$:

$$\bar{x} = \frac{\sum n_i t_i}{n} = \frac{100 \cdot 2 + 170 \cdot 5 + 240 \cdot 10 + 310 \cdot 2 + 380 \cdot 1}{20} = \frac{4450}{20} = 222{,}5.$$

Визначимо D_B:

$$D_B = \frac{100^2 \cdot 2 + 170^2 \cdot 5 + 240^2 \cdot 10 + 310^2 \cdot 2 + 380^2 \cdot 1}{20} - 222{,}5^2 = \frac{1077100}{20} - 49506{,}25 =$$
$$53855 - 49506{,}25 = 4348{,}75.$$

Виправлене середнє квадратичне відхилення дорівнюватиме:

$$S = \sqrt{\frac{n}{n-1} D_B} = \sqrt{\frac{20}{20-1} \cdot 4348{,}75} \approx 67{,}66.$$

За таблицею значень $\int_0^t f(t)dt = \gamma = 0{,}99$ розподілу Стьюдента за заданою надійністю $\gamma = 0{,}99$ і числом ступенів свободи $k = n - 1 = 20 - 1 = 19$ знаходимо значення $t_\gamma(\gamma = 0{,}99; k = 19) = 2{,}86$.

Обчислимо кінці довірчого інтервалу:

$$\bar{x} - \frac{t_\gamma \cdot S}{\sqrt{n}} = 222{,}5 - \frac{2{,}86 \cdot 67{,}66}{\sqrt{0}} = 222{,}5 - 43{,}27 = 179{,}23 \text{ год.};$$

$$\bar{x} + \frac{t_\gamma \cdot S}{\sqrt{n}} = 222{,}5 + \frac{2{,}86 \cdot 67{,}66}{\sqrt{0}} = 222{,}5 + 43{,}27 = 265{,}77 \text{ год.}$$

Отже, з надійністю $\gamma = 0{,}99$ можна стверджувати, що $\overline{X_\Gamma} = a$ буде міститися в інтервалі

$$179{,}23 < a < 265{,}77.$$

При великих обсягах вибірки, а саме: $n > 30$ на підставі центральної граничної теореми теорії ймовірностей (теореми Ляпунова) розподіл Стьюдента наближається до нормального закону. У цьому разі t_γ знаходиться за таблицею значень функції Лапласа.

9.3.3. Побудова довірчих інтервалів із заданою надійністю γ для D_Γ, σ_Γ

У разі, коли ознака X має нормальний закон розподілу, для побудови довірчого інтервалу із заданою надійністю γ для D_Γ, σ_Γ застосовуємо випадкову величину

$$\chi^2 = \frac{n-1}{\sigma_\Gamma^2} S^2, \tag{9.21}$$

що має розподіл χ^2 із $= n - 1$ ступенями свободи.

Оскільки випадкові події

$$A(\chi_1^2 < \chi^2 < \chi_2^2) \text{ і } B\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right)$$

є рівноймовірними, тобто їх імовірності рівні $(P(A) = P(B))$, то маємо:

$$P(\chi_1^2 < \chi^2 < \chi_2^2) = P\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right). \tag{9.22}$$

Підставляючи в (9.22) $\chi^2 = \frac{n-1}{\sigma_\Gamma^2} S^2$ дістанемо

$$P\left(\frac{1}{\chi_2^2} < \frac{1}{\chi^2} < \frac{1}{\chi_1^2}\right) = P\left(\frac{1}{\chi_2^2} < \frac{1}{\frac{n-1}{\sigma_\Gamma^2}S^2} < \frac{1}{\chi_1^2}\right) = P\left(\frac{1}{\chi_2^2} < \frac{\sigma_\Gamma^2}{(n-1)S^2} < \frac{1}{\chi_1^2}\right) =$$

$$= P\left(\frac{(n-1)S^2}{\chi_2^2} < \sigma_\Gamma^2 < \frac{(n-1)S^2}{\chi_1^2}\right) = \gamma.$$

Отже, довірчий інтервал для $\sigma_\Gamma^2 = D_\Gamma$ матиме вигляд:

$$\frac{(n-1)S^2}{\chi_2^2} < D_\Gamma < \frac{(n-1)S^2}{\chi_1^2}. \qquad (9.23)$$

Тоді довірчий інтервал для σ_Γ випливає із (9.23) і буде таким:

$$\frac{S\sqrt{n-1}}{\chi_2} < \sigma_\Gamma < \frac{S\sqrt{n-1}}{\chi_1}. \qquad (9.24)$$

Значення χ_1^2, χ_2^2 знаходимо за таблицею згідно з рівностями:

$$P(\chi^2 > \chi_1^2) = 1 - \frac{\alpha}{2}; \qquad (9.25)$$

$$P(\chi^2 > \chi_2^2) = \frac{\alpha}{2}; \qquad (9.26)$$

$$\alpha = 1 - \gamma.$$

Приклад 9.8. У результаті статистичних досліджень випадкової величини X отримано вибірку об'єму $n = 25$ із таким статистичним розподілом:

x_i	0	1	2	3	4	5	6
n_i	1	3	4	6	5	4	2

З надійністю $\gamma = 0{,}99$ побудувати довірчі інтервали для D_Γ, σ_Γ.

Розв'язання. Для побудови довірчих інтервалів необхідно знайти значення S^2, S.

Обчислимо значення $\bar{x}$:

$$\bar{x} = \frac{\sum x_i n_i}{n} = \frac{0\cdot1+1\cdot3+2\cdot4+3\cdot6+4\cdot5+5\cdot4+6\cdot2}{25} = \frac{81}{25} = 3{,}24.$$

Обчислимо D_B:

$$D_B = \frac{0^2\cdot 1+1^2\cdot 3+2^2\cdot 4+3^2\cdot 6+4^2\cdot 5+5^2\cdot 4+6^2\cdot 2}{25} - 3{,}24^2 = \frac{325}{25} - 10{,}498 =$$
$$= 13 - 10{,}498 = 2{,}502.$$

Виправлена дисперсія і виправлене середнє квадратичне відхилення дорівнюватимуть:

$$S^2 = \frac{n}{n-1}\cdot D_B = \frac{25}{25-1}\cdot 2{,}502 = 2{,}634;$$

$$S = \sqrt{S^2} = \sqrt{2{,}634} \approx 1{,}623.$$

Оскільки $\alpha = 1 - \gamma = 1 - 0{,}99 = 0{,}01$, то згідно з (9.25), (9.26) знаходимо значення χ_1^2, χ_2^2, а саме:

$$P(\chi^2 > \chi_1^2) = 1 - \frac{0{,}01}{2} = 0{,}995;$$

$$P(\chi^2 > \chi_2^2) = \frac{0{,}01}{2} = 0{,}005.$$

За таблицею знаходимо:

$$\chi_1^2(0{,}995; 24) = 10{,}9;$$

$$\chi_2^2(0{,}005; 24) = 45{,}5.$$

Обчислимо кінці довірчого інтервалу для $D_Г$:

$$\frac{(n-1)S^2}{\chi_2^2} = \frac{25-1}{45{,}5}\cdot 2{,}634 = 1{,}389;$$
$$\frac{(n-1)S^2}{\chi_1^2} = \frac{25-1}{10{,}9}\cdot 2{,}634 = 5{,}799.$$

Отже, довірчий інтервал для $D_Г$ буде таким:

$$1{,}389 < D_Г < 5{,}799.$$

Довірчий інтервал для $\sigma_Г$ становить

$$1{,}179 < \sigma_Г < 2{,}408.$$