ЛЕКЦІЯ 14
2. ОСНОВНІ РОЗПОДІЛИ

2.1. Розподіл χ^2

Означення 8. Нехай $x_1, x_2, ..., x_n$ - незалежні нормально розподілені ВВ з параметрами (0;1). Тоді ВВ $\chi_n^2 = \sum_{i=1}^{n} x_i^2$ має розподіл χ^2 з n степенями свободи, що позначається як $\chi^2(n)$.

Щільність ймовірності:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, x \geq 0 \\ 0, x < 0 \end{cases}, \qquad \text{де} \quad \Gamma(m) = \int_0^{+\infty} y^{m-1} e^{-y} dy \quad \text{- гамма-функція.}$$

Числові характеристики.

$$m(\chi^2) = n$$

$$D(\chi^2) = 2n$$

$$A_s = \frac{2^{\frac{3}{2}}}{\sqrt{n}}$$

$$E_k = \frac{12}{n}$$

Застосовується: при побудові довірчих інтервалів і перевірці статистичних гіпотез.

2.2. Розподіл Стьюдента

Означення 9. Нехай ξ і $\chi_n{}^2$ – незалежні випадкові величини, причому $\chi_n^2 \sim \chi^2(n)$. Тоді ВВ $t_n = \dfrac{\xi}{\sqrt{\dfrac{n}{\chi_n^2}}}$ має розподіл Стьюдента (t–розподіл) з n ступенями свободи, що позначається як $t_n \sim S(n)$.

Щільність ймовірності:

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}$$

Зауваження 6. Можна показати, що при $n \to \infty$ щільність ймовірності ВВ $t_n \approx S(n)$ збігається до щільності стандартного нормального розподілу $N(0;1)$, тобто

$$f_{t_n}(x) \to \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad n \to +\infty$$

При $n > 30$ розподіл Стьюдента практично не відрізняється від $N(0;1)$

$$m(t_n) = 0$$

$$D(t_n) = \frac{n}{n-2}, \ (n > 2)$$

$A_S = 0$

$E_K = \dfrac{6}{n-4}$

Застосовується: При побудові довірчих інтервалів і перевірці статистичних гіпотез.

2.3. Розподіл Фішера

Означення 10. Нехай χ_n^2 і χ_m^2 має розподіл χ^2 з n та m ступенями свободи відповідно. Тоді ВВ $F_{n,m} = \dfrac{\chi_n^2}{\chi_m^2} \cdot \dfrac{m}{n}$ має розподіл Фішера (F-розподіл) з n та m ступенями свободи, що записується як $F_{n,m} \sim F(n,m)$.

Щільність ймовірності.

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\dfrac{n+m}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)\Gamma\left(\dfrac{m}{2}\right)} n^{\frac{n}{2}} m^{\frac{m}{2}} \dfrac{x^{\frac{n}{2}-1}}{(m+nx)^{\frac{n+m}{2}}}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

Числові характеристики:

$$m(F) = \frac{m}{m-2}, (m > 2)$$

$$D(F) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)}$$

$$A_S = \frac{(2n+m-2)\sqrt{8(m-4)}}{(m-6)\sqrt{n+m-2}}, (m > 6)$$

3.2. СТАТИСТИЧНІ ОЦІНКИ ПАРАМЕТРІВ РОЗПОДІЛУ

3.2.1. Точкові оцінки параметрів та їх властивості

По обмеженому статистичному матеріалу вибірки досить складно знайти невідомий закон розподілу досліджуваної ознаки генеральної сукупності, яка ототожнюється з випадковою величиною X. Проте з цього обмеженого матеріалу можна одержати досить важливі відомості про випадкову величину, зокрема, оцінити принаймні наближено її числові характеристики – математичне сподівання, дисперсію. Крім того, якщо вид закону розподілу хоча б припустимо відомий, то за вибіркою можна оцінити його параметри, зокрема, a і σ для нормального закону, λ для розподілу Пуассона тощо.

З цією метою застосовуються точкові статистичні оцінки.

Нехай X — випадкова ознака генеральної сукупності з функцією розподілу $F(x,\Theta)$, де Θ — параметр розподілу, числове значення якого невідоме. Припустимо, що з генеральної сукупності зроблена вибірка об'єму n і за цією вибіркою знайдена статистична оцінка Θ^* параметра Θ. Природно припустити, що ця оцінка залежить від вибірки, тобто, якщо зроблено k різних вибірок об'єму n, то буде одержано k статистичних оцінок $\Theta_1^*, \Theta_2^*, ..., \Theta_k^*$ параметра Θ, які в загальному випадку відрізняються одна від одної. Таким чином, статистична оцінка Θ^*

невідомого параметра Θ є величиною випадковою і являється функцією від вибірки $x_1, x_2, ..., x_n$.

Оцінки, які в кожній вибірці визначаються одним певним числом, називаються *точковими* і характеризуються трьома властивостями: незміщеністю, ефективністю і спроможністю.

Означення 3.5. Точкова оцінка Θ^* параметра Θ називається *незміщеною*, якщо її математичне сподівання дорівнює оцінюваному параметру:

$$M(\Theta^*) = \Theta. \quad (3.8)$$

В протилежному випадку оцінка Θ^* називається *зміщеною*. Наприклад, вимірювання фізичної величини приладом, у якого не виставлене нульове значення, дає зміщену оцінку Θ^* справжнього значення Θ цієї величини.

Проте виконання умови (3.8) не дає підстав вважати незміщену оцінку достатньо точним наближенням параметра Θ. Якщо можливі значення Θ^* істотно розсіяні навколо свого середнього значення $M(\Theta^*)$, тобто мають велику дисперсію $D(\Theta^*)$, то знайдена за однією вибіркою оцінка Θ^* може суттєво відрізнятися від оцінюваного параметра Θ. Тому важливу роль відіграє друга властивість точкової оцінки – її ефективність.

Означення 3.6. Точкова оцінка Θ^* параметра Θ називається *ефективною*, якщо дисперсія $D(\Theta^*)$ мінімальна при заданому обсязі вибірки.

Наприклад, ефективність вимірювання фізичної величини приладом буде тим вищою, чим менша вказана в технічному паспорті приладу допустима похибка вимірювань.

Означення 3.7. Точкова оцінка Θ^* параметра Θ називається *спроможною*, якщо при зростанні об'єму вибірки n вона прямує за ймовірністю до оцінюваного параметра:

$$\lim_{n\to\infty} P\left\{\left|\Theta - \Theta^*\right| < \varepsilon\right\} = 1 \quad \text{при довільному } \varepsilon > 0.$$

Наприклад, частота появи події в схемі Бернуллі є спроможною оцінкою ймовірності p появи події в кожному випробуванні, що випливає з теореми Бернуллі (п.2.6.3).

3.2.2. Точкові оцінки математичного сподівання і дисперсії

Якщо для дослідження ознаки X генеральної сукупності проведена вибірка об'єму n і за нею складені дискретний і інтервальний статистичний розподіли, то за оцінку математичного сподівання $M(X)$ ознаки X приймається середнє арифметичне $\overline{x}_B$ варіант вибірки

$$\overline{x}_B = \frac{1}{n}\sum_{i=1}^{k} x_i n_i, \quad (3.9)$$

де x_i – варіанти в дискретному розподілі або центри інтервалів в інтервальному розподілі, n_i – відповідні частоти.

Число $\overline{x}_B$ називається емпіричним або вибірковим математичним сподіванням або *вибірковою середньою* $M(X) \approx \overline{x}_B$. Ця оцінка є незміщеною, спроможною і ефективною.

По аналогії з математичним сподіванням за точкову оцінку для дисперсії $D(X)$ випадкової ознаки X природно прийняти величину

$$D_B = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_B)^2 n_i \,. \tag{3.18}$$

яка називається *вибірковою* або *емпіричною* дисперсією.

Проте вибіркова дисперсія D_B є зміщеною оцінкою для $D(X)$, тобто для неї $M(D_B) \neq D(X)$.

Для виконання умови незміщеності (3.8) за оцінку для дисперсії $D(X)$ обирається величина, яка позначається s^2:

$$s^2 = \frac{n}{n-1}D_B = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x}_B)^2 n_i \,, \tag{3.23}$$

і називається *виправленою* вибірковою дисперсією.

Оцінки D_B і s^2 є спроможними та ефективними для дисперсії $D(X)$.

За точкову оцінку середнього квадратичного відхилення $\sigma(X)$ приймають вибіркове середнє квадратичне відхилення $\sigma_B = \sqrt{D_B}$ або виправлене вибіркове середнє квадратичне відхилення $s = \sqrt{s^2}$.

3.2.3. Поняття інтервальної оцінки. Довірча ймовірність та довірчий інтервал

Використання точкових оцінок параметрів розподілу має два недоліки:

– при малому об’ємі вибірки точкова оцінка може дати велику похибку;

– імовірність того, що випадкова величина (в даному випадку оцінка Θ^*) прийме значення, точно рівне оцінюваному параметру Θ, для дискретних випадкових величин мала, а для неперервних – взагалі рівна нулю.

Тому для оцінки параметра Θ застосовується деякий інтервал $(\Theta^* - \delta;\ \Theta^* + \delta)$, який з досить великою ймовірністю γ покриває цей параметр, тобто

$$P\{\Theta^* - \delta < \Theta < \Theta^* + \delta\} = P\{|\,\Theta - \Theta^*\,| < \delta\} = \gamma \,.$$

Такий інтервал називається *довірчим інтервалом* для параметра Θ, а число γ — *довірчою ймовірністю* або *надійністю* знайденої оцінки. Надійність γ обирається, як правило, рівною 0,95, 0,99 або 0,999. При таких значеннях γ покривання параметра Θ довірчим інтервалом є практично достовірною подією. Величина δ визначає точність оцінки.

Як частинні приклади розглянемо побудову довірчого інтервалу для невідомого параметра a випадкової ознаки X, яка в генеральній сукупності має нормальний розподіл з відомим або невідомим параметром σ.

3.2.4. Довірчий інтервал для математичного сподівання a нормально розподіленої ознаки X при відомому σ

Оцінкою математичного сподівання a є вибіркова середня $\overline{x}_B$, тому за довірчий інтервал для a вибирається інтервал $(\overline{x}_B - \delta;\ \overline{x}_B + \delta)$, для якого виконується умова

$$P\{\overline{x}_B - \delta < a < \overline{x}_B + \delta\} = P\{|\,\overline{x}_B - a\,| < \delta\} = \gamma \,. \tag{3.24}$$

Для визначення δ розглянемо наступну теорему.

<u>Теорема 3.1.</u> Якщо випадкова величина X підпорядкована нормальному закону розподілу з параметрами a і σ, а $x_1, x_2, ..., x_n$ — ряд незалежних спостережень над величиною X, кожне з яких має ті ж числові характеристики, що і X, тобто $M(x_i)=a,\ D(x_i)=\sigma^2$, то вибіркова середня $\overline{x}_B$ також має нормальний розподіл з параметрами a і $\sigma/\sqrt{n}$.

Для нормально розподіленої вибіркової середньої $\overline{x}_B$ з урахуванням $\sigma(\overline{x}_B)=\sigma/\sqrt{n}$ одержимо

$$P\{|\,\overline{x}_B-a\,|<\delta\}=2\Phi\left(\frac{\sqrt{n}\delta}{\sigma}\right), \tag{3.26}$$

а з порівняння формул (3.24) і (3.26) маємо

$$2\Phi\left(\frac{\sqrt{n}\delta}{\sigma}\right)=\gamma.$$

Позначимо $\dfrac{\sqrt{n}\delta}{\sigma}=t,$ тоді $2\Phi(t)=\gamma$.

Задаючи значення надійності γ, за таблицею для функції $\Phi(x)$ (додаток 2) знайдемо відповідне значення t, а з формули

$$\delta=\frac{\sigma t}{\sqrt{n}}. \tag{3.28}$$

— значення δ.

Формула (3.24) набуває вигляду

$$P\{\overline{x}_B-\frac{\sigma t}{\sqrt{n}}<a<\overline{x}_B+\frac{\sigma t}{\sqrt{n}}\}=\gamma \tag{3.29}$$

і задає довірчий інтервал, який з заданою надійністю γ покриває невідомий параметр a нормально розподіленої ознаки X генеральної сукупності.

3.2.5. Довірчий інтервал для математичного сподівання нормально розподіленої ознаки при невідомому σ

Нехай ознака X генеральної сукупності має нормальний розподіл з невідомими параметрами a і σ, і з генеральної сукупності зроблена вибірка об'єму n і обчислені вибіркова середня $\overline{x}_B$ і виправлена вибіркова дисперсія s^2, які є випадковими величинами.

Для побудови довірчого інтервалу, який з заданою надійністю γ покриває невідоме математичне сподівання a ознаки X, в цьому випадку розглядається випадкова величина

$$t=\frac{\overline{x}_B-a}{s/\sqrt{n}},$$

яка має **розподіл Стьюдента** з $(n-1)$ степенями свободи.

За заданою довірчою ймовірністю γ і об'ємом вибірки n по таблиці (додаток 4), складеній за таблицею розподілу Стьюдента, знаходиться значення t_γ, для якого виконується умова

$$P\left\{\left|\frac{\overline{x}_B-a}{s/\sqrt{n}}\right|<t_\gamma\right\}=\gamma$$

або

$$P\{\overline{x}_B-t_\gamma\frac{s}{\sqrt{n}}<a<\overline{x}_B+t_\gamma\frac{s}{\sqrt{n}}\}=\gamma. \tag{3.30}$$

Одержана умова визначає довірчий інтервал, який з заданою надійністю γ покриває невідоме математичне сподівання a ознаки X генеральної сукупності.

3.2.6. Довірчий інтервал для дисперсії σ^2 нормально розподіленої ознаки

Якщо ознака X генеральної сукупності має нормальний розподіл з невідомою дисперсією σ^2 і сформована вибірка $x_1, x_2, ..., x_n$ об'єму n, де варіанти x_i мають той же розподіл, що і X, то для побудови довірчого інтервалу для σ^2 розглядається випадкова величина

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}, \tag{3.31}$$

яка має **розподіл хі-квадрат** з $n-1$ степенями свободи.

Для заданої великої надійності γ вибираються такі два значення χ_1^2 і χ_2^2, щоб виконувалась умова

$$P\left\{\chi_1^2 < \frac{(n-1)s^2}{\sigma^2} < \chi_2^2\right\} = \gamma$$

або для оцінки σ^2 – відповідно умова

$$P\left\{\frac{(n-1)s^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)s^2}{\chi_1^2}\right\} = \gamma. \tag{3.32}$$

Значення χ_1^2 і χ_2^2 вибираються, як правило, такими, що задовольняють умовам

$$P\{\chi^2 < \chi_1^2\} = P\{\chi^2 > \chi_2^2\} = \frac{1-\gamma}{2}, \tag{3.33 }$$

тобто, щоб імовірність виходу випадкової величини (3.31) за межі інтервалу $(\chi_1^2; \chi_2^2)$ була незначною, оскільки надійність γ вибирається близькою до одиниці. Якщо $F(x)$ функція розподілу величини (3.31), то з умов (3.33) випливає

$$F(\chi_1^2) = P\{\chi^2 < \chi_1^2\} = \frac{1-\gamma}{2},$$

$$F(\chi_2^2) = 1 - P\{\chi^2 > \chi_2^2\} = 1 - \frac{1-\gamma}{2} = \frac{1+\gamma}{2}.$$

Величини χ_1^2 і χ_2^2 вибираються з таблиці розподілу χ^2 (додаток 3) за відомими $(n-1)$ степенями свободи і значеннями $\alpha_1 = \frac{1+\gamma}{2}$ і $\alpha_2 = \frac{1-\gamma}{2}$, а за формулою (3.32) знаходиться довірчий інтервал для дисперсії σ^2:

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

або для середнього квадратичного відхилення σ:

$$s\sqrt{\frac{(n-1)}{\chi_2^2}} < \sigma < s\sqrt{\frac{(n-1)}{\chi_1^2}}. \tag{3.34}$$

3.2.6. Методи точкового оцінювання

3.2.6.1. Метод моментів

Нехай $x_1, x_2, ..., x_n$ – вибірка з розподілу $F(x, \theta_1, \theta_2, ..., \theta_s)$. Необхідно одержати оцінки для невідомих параметрів $\theta_1, \theta_2, ..., \theta_s$. Першим загальним методом оцінки є метод моментів, запропоновані К. Пірсоном.

Суть методу полягає в прирівнюванні певної кількості вибіркових моментів $\tilde{m}_k$ відповідним теоретичним моментам $m_k = \int_{-\infty}^{\infty} x^k f(x, \theta_1, \theta_2, ..., \theta_5) dx$

Розглянемо кількість моментів, що дорівнює кількості невідомих параметрів та одержимо систему рівнянь
$$\begin{cases} \tilde{m}_1 = \int_{-\infty}^{\infty} x f(x, \theta_1, \theta_2, ..., \theta_n) dx \\ \tilde{m}_2 = \int_{-\infty}^{\infty} x f(x, \theta_1, \theta_2, ..., \theta_n) dx \\ \quad ... \\ \tilde{m}_k = \int_{-\infty}^{\infty} x f(x, \theta_1, \theta_2, ..., \theta_n) dx \end{cases}$$

3.2.6.2. Метод максимальної правдоподібності

Метод запропонований Фішером:

<u>Дискретні ВВ.</u> Нехай $x_1, x_2, ..., x_n$ - вибірка з дискретної випадкової величини X із заданим законом розподілу $F(x, \theta)$. Необхідно оцінити невідомий параметр θ.

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

<u>Означення 16.</u> Функцією правдоподібності дискретної випадкової величини X називається функція аргументу θ:

$L(x_1; x_2; ...; x_n; \theta) = p(x_1; \theta) p(x_2; \theta) ... p(x_n; \theta)$, де $x_1, x_2, ..., x_n$ - фіксовані числа.

<u>Означення 17</u>. Оцінка $\theta^* = \theta^*(x_1; x_2; ...; x_n)$ знайдену за умови максимуму функції правдоподібності, тобто $L(x_1; x_2; ...; x_n; \theta) \to \max$ називається оцінкою максимальної правдоподібності.

Функції L і $\ln L$ досягають максимуму при одному і тому ж значенні, тому зручніше шукати $\max$ функції $\ln L$.

<u>Означення 18</u>. Логарифмічною функцією правдоподібності називається функція $\ln L = \ln p(x_1; \theta) + \ln p(x_2; \theta) + ... + \ln p(x_n; \theta)$

Етапи пошуку $\max \ln L$:

1.) Знаходимо $\dfrac{d \ln L}{d\Theta}$;

2.) Знаходимо критичну точку θ^* з розв'язку рівняння: $\dfrac{d \ln L}{d\theta} = 0$;

3.) Знаходимо $\dfrac{d^2 \ln L}{d\theta^2}$. Якщо $\dfrac{d^2 \ln L}{d\theta^2} < 0$ в точці θ^*, то θ^* - точка $\max$.