

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
імені ІГОРЯ СІКОРСЬКОГО»

# **МАТЕМАТИЧНА СТАТИСТИКА: ЗБІРНИК ЗАДАЧ**

*Рекомендовано Методичною радою КПІ ім. Ігоря Сікорського  
як навчальний посібник для студентів,  
які навчаються за спеціальністю 124 «Системний аналіз»*

Київ  
КПІ ім. Ігоря Сікорського  
2019

Математична статистика: Збірник задач [Електронний ресурс] : навч. посіб. для студ. спеціальності 124 «Системний аналіз», / КПІ ім. Ігоря Сікорського ; уклад.: І. Ю. Каніовська, О. В. Стусь. – Електронні текстові дані (1 файл: 640 Кбайт). – Київ : КПІ ім. Ігоря Сікорського, 2019. – 124 с.

*Гриф надано Методичною радою КПІ ім. Ігоря Сікорського (протокол № 8 від 25.04.2019 р.)  
за поданням Вченої ради ІПСА (протокол № 4 від 22.04.2019 р.)*

Електронне мережне навчальне видання

# МАТЕМАТИЧНА СТАТИСТИКА: ЗБІРНИК ЗАДАЧ

Укладачі: *Каніовська Ірина Юріївна*, канд. фіз.-мат. наук, доц.  
*Стусь Олександр Вікторович*, канд. фіз.-мат. наук.

Відповідальний редактор *Бондаренко В. Г.*, д-р фіз.-мат. наук, проф.

Рецензент: *Льєнко А. Б.*, канд. фіз.-мат. наук, доц.

Посібник містить задачі для практичних занять та самостійного розв'язання по основних розділах дисципліни «Математична статистика» – точкове оцінювання параметрів розподілу генеральної сукупності, методи побудови точкових оцінок, інтервальне оцінювання параметрів закону розподілу, перевірка статистичних гіпотез. Кожен розділ містить методичні вказівки із прикладами розв'язання. Також наведено відповіді до запропонованих задач.

Для студентів математичних і технічних спеціальностей університетів.

# Зміст

Передмова.....	5
Глава 1. Точкове оцінювання невідомих параметрів розподілу генеральної сукупності.....	6
1.1. Точкові оцінки.....	6
1.2. Оцінки з мінімальною дисперсією.....	16
Задачі для самостійного розв'язання.....	21
Глава 2. Методи побудови точкових оцінок.....	33
2.1. Емпіричні оцінки.....	33
2.2. Метод моментів.....	38
2.3. Метод максимальної правдоподібності.....	41
Задачі для самостійного розв'язання.....	45
Глава 3. Побудова довірчих інтервалів для параметрів закону розподілу.....	52
Задачі для самостійного розв'язання.....	62
Глава 4. Критерій Пірсона ( $\chi^2$ ).....	68
4.1. Перевірка гіпотези про вид розподілу генеральної сукупності за критерієм $\chi^2$ .....	69
4.2. Застосування критерію $\chi^2$ в схемі Бернуллі.....	75
4.3. Критерій незалежності $\chi^2$ .....	77
Задачі для самостійного розв'язання.....	79
Глава 5. Перевірка гіпотез про параметри розподілу генеральної сукупності.....	86

5.1. Перевірка простих гіпотез про параметри розподілу генеральної сукупності .....	86
5.2. Перевірка складних гіпотез про параметри розподілу генеральної сукупності .....	93
Задачі для самостійного розв'язання .....	104
Відповіді .....	112
Список літератури .....	116
Додатки. Таблиці математичної статистики .....	117

## Передмова

Математична статистика є розділом математики, який займається розробкою методів отримання науково обґрунтованих висновків про масові випадкові явища по даним спостережень та експериментів.

Математична статистика основана на поняттях і методах теорії ймовірностей. Задачі математичної статистики в деякій мірі є оберненими до задач теорії ймовірностей. Якщо в теорії ймовірностей вважається заданою ймовірнісна модель випадкового явища, згідно якої робляться розрахунки ймовірностей подій, що нас цікавлять, то в математичній статистиці вважається, що ймовірнісна модель випадкового явища не задана (або задана частково). На основі отриманих статистичних даних підбирається ймовірнісна модель для отримання належних висновків.

Метою навчального посібника є надання допомоги студентам в опануванні методики розв'язання основних задач з курсу математичної статистики. Зміст глав відповідає основним темам практичних занять з математичної статистики, що викладається для студентів Інституту прикладного системного аналізу НТУУ «КПІ ім. Ігоря Сікорського». Автори не внесли в посібник матеріал первинної обробки статистичних даних, оскільки існує для цього багато пакетів комп'ютерних програм. Наприклад, STASTISTICA, SYSTAT, STATGRAPHICS, STATA тощо. На початку кожного розділу вміщено короткі теоретичні відомості та формули, наведено зразки розв'язання типових задач. Всі задачі, які зібрані в посібнику, мають загальний характер, тому його можна використовувати на різних факультетах вищих навчальних закладів, де вивчають методи математичної статистики.

# Глава 1. Точкове оцінювання невідомих параметрів розподілу генеральної сукупності

## 1.1. Точкові оцінки

Точковою оцінкою  $\theta^*$  параметра  $\theta$  називають будь-яку функцію результатів спостережень над випадковою величиною  $\xi$  (статистику), за допомогою якої судять про значення параметра  $\theta$ :

$$\theta^* = \theta^*(\xi_1, \xi_2, \dots, \xi_n).$$

Зрозуміло, що точкова оцінка  $\theta^*$ , на відміну від параметра  $\theta$ , є випадковою величиною, яка залежить від закону розподілу випадкової величини  $\xi$  та обсягу вибірки. Про якість оцінки можна судити не по її конкретним значенням, а лише по вибіркового розподілу оцінки. У зв'язку з цим розглянемо вимоги, що висувають до оцінок.

Точкова оцінка  $\theta^*$  параметра  $\theta$  називається *незміщеною*, якщо її математичне сподівання дорівнює оцінюваному параметру  $\theta$ , тобто  $E\theta^* = \theta$ . Якщо ця умова не виконується, то оцінка  $\theta^*$ , яка отримана за різними вибірками, буде в середньому або завищувати значення параметра  $\theta$ , або це значення занижувати. Якщо  $E\theta^* = \theta + a(\theta)$ , то величину  $a(\theta)$  називають *зсувом точкової оцінки*.

Точкова оцінка  $\theta^*$  параметра  $\theta$  називається *асимптотично незміщеною*, якщо  $E\theta^* \rightarrow \theta$  при  $n \rightarrow \infty$ .

Послідовність точкових оцінок  $\theta_n^* = \theta_n^*(\xi_1, \xi_2, \dots, \xi_n)$ ,  $n \in \mathbb{N}$  параметра розподілу  $\theta$  називається *конзистентною (спроможною)*, якщо

задовольняє закону великих чисел, тобто збігається за ймовірністю до параметра  $\theta$  при  $n \rightarrow \infty$ :

$$\lim_{n \rightarrow \infty} P\{|\theta_n^* - \theta| < \varepsilon\} = 1.$$

У випадку використання конзистентних оцінок виправдано використання вибірок великого обсягу, тому що при цьому стають малоймовірними значні помилки при оцінюванні. Отже, практичний зміст мають лише конзистентні оцінки.

Наведемо простий достатній **критерій конзистентності незміщених точкових оцінок**. Нехай точкова оцінка  $\theta_n^* = \theta_n^*(\xi_1, \xi_2, \dots, \xi_n)$  є незміщеною оцінкою параметра  $\theta$ , з нерівності Чебишева випливає, що, якщо  $\lim_{n \rightarrow \infty} D\theta_n^* = 0$ , то незміщена точкова оцінка буде конзистентною. Аналогічний критерій є справедливим і для асимптотично незміщених точкових оцінок.

### **Приклади розв'язання задач**

**Приклад 1.1.1.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & \text{якщо } x \in [a, b]; \\ 0, & \text{якщо } x \notin [a, b]. \end{cases}$$

Розглядаються оцінки:

$$\theta_1^* = \min\{\xi_1, \xi_2, \dots, \xi_n\}; \theta_2^* = \max\{\xi_1, \xi_2, \dots, \xi_n\};$$

$$\theta_3^* = \frac{1}{n} \sum_{i=1}^n \xi_i; \theta_4^* = \frac{\xi_{n-1} + \xi_n}{2}.$$

Які з оцінок  $\theta_1^*, \theta_2^*, \theta_3^*, \theta_4^*$  та яких параметрів (можливо відмінних від  $a$  і  $b$ ) є незміщеними оцінками? Конзистентними оцінками?

**Розв'язання.** Розглянемо кожну з наведених оцінок.

Оцінка  $\theta_2^* = \max\{\xi_1, \xi_2, \dots, \xi_n\}$ . Знайдемо розподіл оцінки  $\theta_2^*$ . За умовою  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з рівномірного на відрізку  $[a, b]$  розподілу. Це означає, що  $\xi_1, \xi_2, \dots, \xi_n$  – незалежні випадкові величини, кожна з яких розподілена рівномірно на відрізку  $[a, b]$ . Тому

$$F_{\theta_2^*}(x) = P\{\theta_2^* < x\} = P\{\max\{\xi_1, \xi_2, \dots, \xi_n\} < x\} = \prod_{i=1}^n P\{\xi_i < x\}.$$

Остання рівність справджується внаслідок незалежності випадкових величин  $\xi_1, \xi_2, \dots, \xi_n$ . Враховуючи, що розподіли випадкових величин  $\xi_1, \xi_2, \dots, \xi_n$  є неперервними зі щільністю  $f(x; a, b)$ , маємо

$$\begin{aligned} F_{\theta_2^*}(x) &= \prod_{i=1}^n P\{\xi_i < x\} = \\ &= \prod_{i=1}^n \int_{-\infty}^x f(y; a, b) dy = \left( \int_{-\infty}^x f(y; a, b) dy \right)^n. \end{aligned}$$

Звідси випливає, що  $\theta_2^*$  – неперервна випадкова величина. Її щільність

$$f_{\theta_2^*}(x) = \frac{d}{dx} F_{\theta_2^*}(x) = n f(x; a, b) \left( \int_{-\infty}^x f(y; a, b) dy \right)^{n-1}.$$

Функція розподілу випадкової величини  $\xi_i$

$$F_{\xi_i}(x) = \begin{cases} 0, & \text{якщо } x < a; \\ \frac{x-a}{b-a}, & \text{якщо } a \leq x \leq b; \\ 1, & \text{якщо } x > b. \end{cases}$$

Отже,

$$f_{\theta_2^*}(x) = \begin{cases} 0, & \text{якщо } x \notin [a, b]; \\ \frac{n(x-a)^{n-1}}{(b-a)^n}, & \text{якщо } x \in [a, b]. \end{cases}$$

За відомою щільністю розподілу оцінки  $\theta_2^*$  можемо обчислити її математичне сподівання:

$$E\theta_2^* = \frac{n}{n+1}b + \frac{a}{n+1} \rightarrow b \text{ при } n \rightarrow \infty.$$

Останнє означає, що  $\theta_2^*$  є асимптотично незміщеною оцінкою параметра  $b$ .

З'ясуємо, чи є  $\theta_2^*$  конзистентною оцінкою параметра  $b$ . З цією метою перевіримо, чи буде  $\theta_2^*$  збігатися за ймовірністю до  $b$ . Якщо  $\varepsilon > b - a$ , то  $P\{|\theta_2^* - b| > \varepsilon\} = 0$ .

При  $\varepsilon \leq b - a$  маємо

$$\begin{aligned} P\{|\theta_2^* - b| > \varepsilon\} &= P\{\theta_2^* \in (-\infty, b - \varepsilon) \cup (b + \varepsilon, +\infty)\} = \\ &= \int_{-\infty}^{b-\varepsilon} f_{\theta_2^*}(x)dx + \int_{b+\varepsilon}^{+\infty} f_{\theta_2^*}(x)dx = \int_a^{b-\varepsilon} f_{\theta_2^*}(x)dx = \\ &= \int_a^{b-\varepsilon} \frac{n(x-a)^{n-1}}{(b-a)^n} dx = \left(1 - \frac{\varepsilon}{b-a}\right)^n. \end{aligned}$$

В обох випадках  $P\{|\theta_2^* - b| > \varepsilon\}$  прямує до 0 при  $n \rightarrow \infty$ . Отже,  $\theta_2^*$  є консистентною оцінкою параметра  $b$ .

Оцінка  $\theta_1^* = \min\{\xi_1, \xi_2, \dots, \xi_n\}$ . Оцінка  $\theta_1^*$  досліджується аналогічно попередній. Розподіл оцінки  $\theta_1^*$  неперервний, а її щільність

$$f_{\theta_1^*}(x) = \begin{cases} 0, & \text{якщо } x \notin [a, b]; \\ \frac{n(b-x)^{n-1}}{(b-a)^n}, & \text{якщо } x \in [a, b]. \end{cases}$$

Звідси

$$E\theta_1^* = \frac{n}{n+1}a + \frac{b}{n+1},$$

тобто  $\theta_1^*$  не є незміщеною оцінкою ні параметра  $a$ , ні параметра  $b$ , але вона є асимптотично незміщеною оцінкою параметра  $a$ .

Як і при дослідженні оцінки  $\theta_1^*$ , встановлюємо, що  $\theta_1^*$  є консистентною оцінкою параметра  $a$ .

Оцінка  $\theta_3^* = \frac{1}{n} \sum_{i=1}^n \xi_i$ . Обчислимо математичне сподівання оцінки  $\theta_3^*$ :

$$E\theta_3^* = E \frac{1}{n} \sum_{i=1}^n \xi_i = \frac{1}{n} \sum_{i=1}^n E\xi_i = \frac{1}{n} n \frac{a+b}{2} = \frac{a+b}{2}.$$

Оскільки  $\xi_i$ ,  $i = \overline{1, n}$  – рівномірно розподілені на відрізку  $[a, b]$  випадкові величини, а для них математичне сподівання  $m_1$ , як відомо, дорівнює  $\frac{a+b}{2}$ , то

$$E\theta_3^* = \frac{a+b}{2} = m_1.$$

Отже,  $\theta_3^*$  не є незміщеною оцінкою ні параметра  $a$ , ні параметра  $b$ , але  $\theta_3^*$  – незміщена оцінка параметра  $m_1 = \frac{a+b}{2}$  – математичного сподівання рівномірного на відрізьку  $[a, b]$  розподілу.

Далі, згідно з законом великих чисел

$$\theta_3^* = \frac{1}{n} \sum_{i=1}^n \xi_i \xrightarrow{P} m_1 = \frac{a+b}{2}, \text{ при } n \rightarrow \infty.$$

тобто  $\theta_3^*$  є конзистентною оцінкою параметра  $m_1$ .

Оцінка  $\theta_4^* = \frac{\xi_{n-1} + \xi_n}{2}$ . Для оцінки  $\theta_4^*$  маємо

$$E\theta_4^* = E \frac{\xi_{n-1} + \xi_n}{2} = \frac{1}{2} (E\xi_{n-1} + E\xi_n) = \frac{a+b}{2}.$$

Отже,  $\theta_4^*$  є незміщеною оцінкою параметра  $m_1 = \frac{a+b}{2}$ .

Оцінка  $\theta_4^*$  не збігається за ймовірністю до жодної константи (параметра). Справді, щільністю розподілу випадкової величини

$\theta_4^* = \frac{\xi_{n-1} + \xi_n}{2}$  є функція

$$f_{\theta_4^*}(x) = \begin{cases} 0, & \text{якщо } x \notin [a, b]; \\ \frac{4(x-a)}{(b-a)^2}, & \text{якщо } x \in \left[ a, \frac{a+b}{2} \right]; \\ \frac{4(b-x)}{(b-a)^2}, & \text{якщо } x \in \left[ \frac{a+b}{2}, b \right]. \end{cases}$$

(Щільність  $f_{\theta_4^*}(x)$  одержано як згортку щільностей рівномірних на відрізьку  $[a, b]$  розподілів, її графік зображено на рис. 1.1). І якщо

припустити, що  $\theta_4^*$  збігається за ймовірністю до деякої константи  $c$ , наприклад, до вказаної на рис. 1.1, то для кожного досить малого  $\varepsilon > 0$

$$P\{|\theta_4^* - c| > \varepsilon\} = \int_{\{x: |x-c| > \varepsilon\}} f_{\theta_4^*}(x) dx$$

є константою, що не залежить від  $n$  (при досить малих  $\varepsilon$  вона дорівнює заштрихованій площі на рис. 1.1) і, отже, не збігається до нуля при  $n \rightarrow \infty$ .

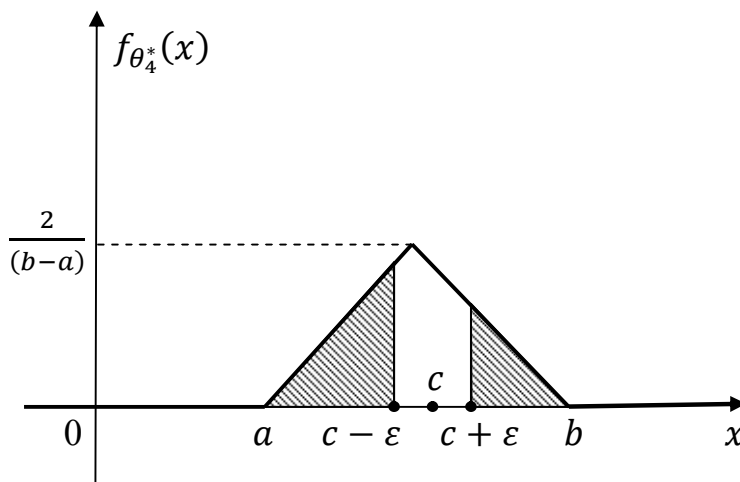


Рисунок 1.1 – Графік щільності розподілу оцінки  $\theta_4^*$

**Приклад 1.1.2.** Нехай маємо випробування у схемі Бернуллі з невідомою ймовірністю успіху  $\theta$ . Розглянемо випадкову вибірку  $\xi_1, \dots, \xi_n$ , де  $\xi_i, i = \overline{1, n}$  – випадкова величина, яка з ймовірністю  $\theta$  приймає значення 1 («успіх» в  $i$ -му випробуванні) і з ймовірністю  $1 - \theta$  – значення 0 («невдача» в  $i$ -му випробуванні).

В якості оцінки  $\theta$  візьмемо відносну частоту успіхів, тобто  $\theta^*(\vec{\xi}) = k(\vec{\xi})/n$ , де  $k(\vec{\xi}) = \sum_{i=1}^n \xi_i$  є сумарна кількість успіхів в  $n$  дослідях. Дослідити дану оцінку на незміщеність та конзистентність для параметра  $\theta$ .

**Розв'язання.** Ця оцінка є незміщеною, оскільки

$$E\theta^*(\vec{\xi}) = \frac{1}{n}Ek(\xi_1, \dots, \xi_n) = \frac{1}{n}\sum_{i=1}^n E\xi_i = \frac{1}{n}n\theta = \theta,$$

і конзистентною, що безпосередньо витікає із закону великих чисел у формі Бернуллі, згідно з яким для будь-якого  $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{k(\vec{\xi})}{n} - p \right| < \varepsilon \right\} = 1.$$

В подальшому статистику  $k(\vec{\xi})$  і її значення будемо позначати  $k$ . У кожному конкретному випадку повинно бути зрозуміло, про що йде мова: про випадкову величину чи її реалізацію.

**Приклад 1.1.3.** Нехай  $\xi_1, \dots, \xi_n$  – випадкова вибірка з генеральної сукупності  $\xi$ , що має нормальний розподіл з невідомим середнім значенням  $\theta$  і відомою дисперсією  $\sigma^2$ . Дослідити оцінку  $\theta^* = \theta^*(\xi_1, \xi_2, \dots, \xi_n) = \xi_1$  на незміщеність та конзистентність для параметра  $\theta$ .

**Розв'язання.** Оцінка  $\theta^*$  є незміщеною для  $\theta$ , так як  $E\xi_1 = E\xi = \theta$ , але не є конзистентною, оскільки, по-перше,  $\xi_1$  не залежить від об'єму вибірки і, відповідно, її розподіл не змінюється з ростом  $n$ , а по-друге,

$$P\{|\xi_1 - \theta| < \varepsilon\} = \frac{2}{\sigma\sqrt{2\pi}} \int_0^\varepsilon e^{-\frac{t^2}{2\sigma^2}} dt \neq 1.$$

**Приклад 1.1.4.** Розглянемо випадкову вибірку  $\xi_1, \dots, \xi_5$  обсягу  $n = 5$  із генеральної сукупності  $\xi$ , розподіленої за показниковим законом

$$f_{\xi}(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{якщо } x \geq 0; \\ 0, & \text{якщо } x < 0. \end{cases}$$

В якості точкової оцінки математичного сподівання візьмемо середнє арифметичне крайніх членів варіаційного ряду

$$\theta^*(\vec{\xi}) = \frac{\xi_{(1)} + \xi_{(5)}}{2}.$$

Перевірити дану оцінку на незміщеність.

**Розв'язання.** Покажемо, що оцінка є зміщеною. Із властивостей математичного сподівання отримаємо

$$E\theta^*(\vec{\xi}_5) = \frac{E\xi_{(1)} + E\xi_{(5)}}{2},$$

звідки робимо висновок, що для обчислення математичного сподівання точкової оцінки необхідно знати закони розподілу випадкових величин  $\xi_{(1)}$  і  $\xi_{(5)}$ . Для першої з них маємо

$$\begin{aligned} F_{\xi_{(1)}}(x) &= P\{\xi_{(1)} < x\} = 1 - P\{\xi_{(1)} \geq x\} = 1 - P\{\xi_1 \geq x, \dots, \xi_5 \geq x\} = \\ &= 1 - P\{\xi_1 \geq x\} \dots P\{\xi_5 \geq x\} = 1 - ((1 - F(x))^5), \end{aligned}$$

де  $F(x)$  – функція розподілу генеральної сукупності.

Аналогічні обчислення для випадкової величини  $\xi_{(5)}$ :

$$F_{\xi_{(5)}}(x) = P\{\xi_{(5)} < x\} = P\{\xi_1 < x, \dots, \xi_5 < x\} = (F(x))^n.$$

Враховуючи вигляд функції розподілу генеральної сукупності (вид показникового закону розподілу), робимо висновок, що

$$F_{\xi_{(1)}}(x) = \begin{cases} 1 - e^{-5\lambda x}, & \text{якщо } x \geq 0; \\ 0, & \text{якщо } x < 0, \end{cases}$$
$$F_{\xi_{(5)}}(x) = \begin{cases} (1 - e^{-5\lambda x})^5, & \text{якщо } x \geq 0; \\ 0, & \text{якщо } x < 0. \end{cases}$$

Отже,

$$E\xi_{(1)} = \int_{-\infty}^{\infty} x f_{\xi_{(1)}}(x) dx = 5\lambda \int_0^{\infty} x e^{-5\lambda x} dx = \frac{1}{5\lambda},$$
$$E\xi_{(5)} = \int_{-\infty}^{\infty} x f_{\xi_{(5)}}(x) dx = 5\lambda \int_0^{\infty} x (1 - e^{-5\lambda x})^4 e^{-\lambda x} dx = \frac{287}{60\lambda}.$$

Із знайдених формул знаходимо

$$E\theta^*(\vec{\xi}_5) = \frac{1}{2} \left( \frac{1}{5\lambda} + \frac{287}{60\lambda} \right) = \frac{299}{120\lambda}.$$

Порівнюючи знайдені значення з математичним сподіванням для розглянутої генеральної сукупності  $\xi$ , переконуємось, що  $E\theta^*(\vec{\xi}) \neq E\xi$ , тобто точкова оцінка  $\theta^*(\vec{\xi})$  зміщена.

## 1.2. Оцінки з мінімальною дисперсією

Незміщену оцінку  $\theta^*$  параметра  $\theta$  називають *ефективною* оцінкою, якщо

$$D\theta^* = \inf_{\theta^*: E\theta^* = \theta} D\theta,$$

тобто має найменшу дисперсію серед усіх можливих незміщених оцінок параметра  $\theta$ , які обчислені за вибірками одного й того ж обсягу  $n$ .

Означення ефективності не дуже сприяє дослідженню оцінки. По-перше, не завжди легко обчислити дисперсію оцінки, по-друге, навіть коли її можна обчислити, то немає гарантії, що вона буде найменшою серед дисперсій усіх інших оцінок, бо не завжди можна розглянути усі оцінки, а тим більше, обчислити їх дисперсію. Але завдання можна полегшити, якщо взяти до уваги нижченаведену нерівність Крамера - Рао.

Нехай  $L(x_1, x_2, \dots, x_n, \theta)$  – функція правдоподібності випадкової величини  $\xi$  з генеральної сукупності  $\Omega_\xi$ , що містить невідомий параметр  $\theta$ .

Функція вигляду

$$I(\theta) = E \left( \frac{\partial(\ln L(\vec{\xi}, \theta))}{\partial \theta} \right)^2 = -E \frac{\partial^2 \ln L(\vec{\xi}, \theta)}{\partial \theta^2}.$$

називається *кількістю інформації за Фішером*. Її називають так тому, що розглядають в деякому розумінні як кількість інформації, що міститься у випадковому векторі  $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$  про значення параметра  $\theta$ . Справедливою буде також рівність

$$I(\xi_1, \xi_2, \dots, \xi_n, \theta) = nI(\xi_i, \theta), \quad i = \overline{1, n}.$$

**Нерівність Крамера - Рао.** Нехай  $\theta^*$  – незміщена оцінка параметра  $\theta$ , тоді

$$D\theta^* \geq \frac{1}{I(\xi_1, \xi_2, \dots, \xi_n, \theta)} = \frac{1}{I((\xi_i, \theta))}. \quad (1.1)$$

Якщо для деякої незміщеної оцінки нерівність Крамера - Рао перетворюється в рівність, то така оцінка й буде ефективною.

Але для перевірки ефективності оцінки необов'язково обчислювати дисперсію оцінки та кількість інформації за Фішером. Можна скористатися **наслідком з нерівності Крамера - Рао.** Незміщена оцінка  $\theta^* = \theta^*(\xi_1, \xi_2, \dots, \xi_n)$  параметра  $\theta$  буде ефективною тоді і тільки тоді, коли

$$\frac{\partial \ln L(\vec{\xi}, \theta)}{\partial \theta} = c(\theta, n)(\theta^* - \theta). \quad (1.2)$$

де функція  $c(\theta, n)$  не залежить від вибірки.

### **Приклади розв'язання задач**

**Приклад 1.2.1.** Розглянемо модель  $N(a, \theta)$ , де середнє значення  $a$  генеральної сукупності відоме, а  $\theta = \sigma^2$  – невідомий параметр.

Довести, що

$$D^* \xi = \frac{1}{n} \sum_{i=1}^n (\xi_i - a)^2$$

є незміщеною і ефективною за Крамером - Рао оцінкою параметра  $\sigma^2$ .

**Розв'язання.** Дійсно,

$$ED^*\xi = \frac{1}{n} \sum_{i=1}^n E(\xi_i - a)^2 = \frac{1}{n} \sum_{i=1}^n D\xi_i = \frac{1}{n} \sum_{i=1}^n \sigma^2 = \frac{1}{n} n\sigma^2 = \sigma^2,$$

тобто  $D^*\xi$  – незміщена оцінка.

Обчислимо дисперсію  $D^*\xi$ :

$$\begin{aligned} D(D^*\xi) &= E(D^*\xi)^2 - (ED^*\xi)^2 = E\left(\frac{1}{n} \sum_{i=1}^n (\xi_i - a)^2\right)^2 - \sigma^4 = \\ &= \frac{1}{n^2} \left( \sum_{i=1}^n E(\xi_i - a)^4 + \sum_{\substack{i,j=1 \\ i \neq j}}^n E((\xi_i - a)^2(\xi_j - a)^2) \right) - \sigma^4 = \\ &= \frac{n\mu_4}{n^2} + \frac{n(n-1)}{n^2} \sigma^4 - \sigma^4 = \frac{3\sigma^4}{n} + \frac{n-1}{n} \sigma^4 - \sigma^4 = \frac{2\sigma^4}{n}. \end{aligned}$$

Потім визначимо кількість інформації за Фішером:

$$\begin{aligned} I(\xi, \theta) &= E\left(\frac{\partial}{\partial \sigma^2} \left( \ln \frac{1}{\sqrt{2\pi}} - \frac{1}{2} \ln \sigma^2 - \frac{(\xi-a)^2}{2\sigma^2} \right)\right)^2 = \\ &= E\left(-\frac{1}{2\sigma^2} + \frac{(\xi-a)^2}{2\sigma^4}\right)^2 = \frac{1}{4\sigma^4} - \frac{E(\xi-a)^2}{2\sigma^6} + \frac{E(\xi-a)^4}{4\sigma^8} = \\ &= \frac{1}{4\sigma^4} - \frac{\sigma^2}{2\sigma^6} + \frac{3\sigma^4}{4\sigma^8} = \frac{1}{2\sigma^4}, \end{aligned}$$

оскільки для нормальної моделі  $\mu_4 = 3\sigma^4$ . В результаті отримаємо, що нерівність (1.1) перетворилася в рівність.

$$DD^*\xi = \frac{1}{nI(\xi, \theta)},$$

тобто  $D^*\xi$  – ефективна оцінка параметра  $\theta$  для нормальної моделі.

**Приклад 1.2.2.** Довести, що оцінка

$$D^{**}\xi = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2$$

для нормальної моделі  $N(a, \theta)$  при невідомому  $a$  не є ефективною.

**Розв'язання.** Відомо, що дана оцінка є незміщеною оцінкою параметра  $\sigma^2$ . Можна показати, що

$$DD^{**}\xi = \frac{2\sigma^4}{n-1} \neq \frac{1}{nI(\xi, \theta)}.$$

тобто задана оцінка є неефективною.

**Приклад 1.2.3.** Розглянемо експоненційну модель

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & \text{якщо } x \geq 0; \\ 0, & \text{якщо } x < 0. \end{cases}$$

Покажемо, що  $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$  є ефективною за Крамером-Рао оцінкою невідомого параметра  $\theta$ .

**Розв'язання.** Перш за все переконаємося у незміщеності заданої оцінки. Як відомо, для показниково розподіленої випадкової величини  $\xi$  з заданою щільністю  $E\xi = \theta$ , а також вибіркове середнє є незміщеною оцінкою для математичного сподівання у випадку його існування.

**Списіб 1.** Обчислимо дисперсію нашої оцінки та кількість інформації за Фішером.

$$\begin{aligned}
D(\bar{\xi}) &= D\left(\frac{1}{n}\sum_{i=1}^n \xi_i\right) = \frac{1}{n}D\xi = \frac{\theta^2}{n}, \\
I(\theta) &= E\left(\frac{\partial}{\partial\theta}\left(\ln\frac{1}{\theta}e^{-\frac{\xi}{\theta}}\right)\right)^2 = E\left(-\frac{1}{\theta} + \frac{\xi}{\theta^2}\right)^2 = \\
&= E\left(\frac{\xi-\theta}{\theta^2}\right)^2 = \frac{E(\xi-\theta)^2}{\theta^4} = \frac{D\xi}{\theta^4} = \frac{\theta^2}{\theta^4} = \frac{1}{\theta^2}.
\end{aligned}$$

Підставляємо в (1.1) та бачимо, що

$$D(\bar{\xi}) = \frac{1}{nI(\bar{\xi},\theta)}.$$

Тобто оцінка  $\bar{\xi}$  є ефективною для параметра  $\theta$ .

**Спосіб 2.** Обчислимо  $\frac{\partial \ln L(\bar{\xi},\theta)}{\partial \theta}$ .

$$\begin{aligned}
\frac{\partial \ln L(\bar{\xi},\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta} \left( \ln \frac{e^{-\frac{\sum_{i=1}^n \xi_i}{\theta}}}{\theta^n} \right) = \frac{\partial}{\partial \theta} \left( -\frac{\sum_{i=1}^n \xi_i}{\theta} - n \ln \theta \right) = \\
&= \frac{\sum_{i=1}^n \xi_i}{\theta^2} - \frac{n}{\theta} = \frac{n}{\theta^2} (\bar{\xi} - \theta).
\end{aligned}$$

Бачимо, що  $\frac{\partial \ln L(\bar{\xi},\theta)}{\partial \theta}$  має вигляд (1.2), що підтверджує ефективність оцінки  $\bar{\xi}$  для параметра  $\theta$ .

## Задачі для самостійного розв'язання

1.1. Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

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

(нормальний розподіл з параметрами  $0; \sigma^2$ ),  $\sigma^2$  належить скінченному відрізку  $[a, b]$ ,  $a > 0$ . Чи є оцінка

$$(\sigma^2)^* = \frac{1}{n} \sum_{i=1}^n \xi_i^2$$

ефективною оцінкою параметра  $\sigma^2$ ?

1.2. Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; a, b) = \begin{cases} \frac{1}{b-a}, & \text{якщо } x \in [a, b]; \\ 0, & \text{якщо } x \notin [a, b]. \end{cases}$$

(рівномірний на відрізку  $[a, b]$  розподіл). Розглядаються оцінки

$$\theta_1^* = \min\{\xi_i\}; \theta_2^* = \max\{\xi_i\}; \theta_3^* = \frac{1}{n} \sum_{i=1}^n \xi_i;$$

$$\theta_4^* = \frac{\xi_n + \xi_{n+1}}{2}; \theta_5^* = \max\{\xi_i\} - \min\{\xi_i\};$$

$$\theta_6^* = \frac{1}{2} (\max\{\xi_i\} + \min\{\xi_i\});$$

$$\theta_7^* = \min\{\xi_i\} - \frac{1}{n-1} (\max\{\xi_i\} - \min\{\xi_i\});$$

$$\theta_8^* = \max\{\xi_i\} + \frac{1}{n-1} (\max\{\xi_i\} - \min\{\xi_i\}).$$

Серед  $\theta_i^*$ ,  $i = \overline{1, 8}$ , визначити незміщені, конзистентні оцінки параметрів  $a$  і  $b$ . Можливо, серед них є незміщені й конзистентні оцінки параметрів, відмінних від  $a$  і  $b$ ?

**1.3.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з пуассонівського розподілу з параметром  $\lambda$ :

$$P(k, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

Позначимо через  $\lambda^*$  точку, в якій функція  $L(\lambda) = \prod_{i=1}^n P(\xi_i; \lambda)$  досягає свого найбільшого значення. Чи є  $\lambda^*$  оцінкою і, якщо так, чи буде вона незміщеною оцінкою параметра  $\lambda$ ? Конзистентною оцінкою  $\lambda$ ?

**1.4.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \sigma^2) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2x}} e^{-\frac{(\ln x - \mu_0)^2}{2\sigma^2}}, & \text{якщо } x > 0; \\ 0, & \text{якщо } x \leq 0. \end{cases}$$

(логарифмічно нормальний розподіл),  $\mu_0$  – відоме. Чи є

$$(\sigma^2)^* = \frac{1}{n} \sum_{i=1}^n (\ln \xi_i - \mu_0)^2$$

ефективною оцінкою параметра  $\sigma^2$ ?

**1.5.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta) = \begin{cases} \frac{1}{2}, & \text{якщо } x \in [\theta - 1, \theta + 1]; \\ 0, & \text{якщо } x \notin [\theta - 1, \theta + 1]. \end{cases}$$

(рівномірний на відрізку  $[\theta - 1, \theta + 1]$  розподіл). Позначимо через  $\theta^*$  точку, в якій функція  $L(\theta) = \prod_{i=1}^n f(\xi_i; \theta)$  досягає свого найбільшого значення. Чи є  $\theta^*$  оцінкою і, якщо так, чи буде вона незміщеною оцінкою параметра  $\theta$ ? Конзистентною оцінкою параметра  $\theta$ ?

**1.6.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \alpha, \theta) = \begin{cases} \frac{1}{\alpha} e^{-\frac{x-\theta}{\alpha}}, & \text{якщо } x \geq \theta; \\ 0, & \text{якщо } x < \theta. \end{cases}$$

(зміщений показниковий розподіл з параметрами  $\alpha$  та  $\theta$ ).

Розглядаються оцінки

$$\theta_1^* = \min\{\xi_i\} - \frac{\bar{\xi} - \min\{\xi_i\}}{n}; \quad \theta_2^* = \bar{\xi} - \theta_1^*.$$

Незміщеними та конзистентними оцінками яких параметрів є  $\theta_1^*$  і  $\theta_2^*$  (можливо, ці параметри відмінні від  $\alpha$  і  $\theta$ )?

**1.7.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(x-\theta)^2}{2\sigma_0^2}}$$

(нормальний розподіл з параметрами  $\theta; \sigma_0^2$ ),  $\sigma_0$  – відоме. Чи є

$$\theta^* = \bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$$

ефективною оцінкою параметра  $\theta$ ?

**1.8.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta) = \begin{cases} 0, & \text{якщо } x < \theta; \\ e^{\theta-x}, & \text{якщо } x \geq \theta. \end{cases}$$

Чи є

$$\theta^* = -\frac{1}{n} + \min \{\xi_i\}$$

незмщеною оцінкою параметра  $\theta$ ? Конзистентною оцінкою  $\theta$ ?

**1.9.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

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

(нормальний розподіл з параметрами  $a; \sigma^2$ ). Позначимо через  $(a^*, (\sigma^2)^*)$  точку, в якій функція  $L(a, \sigma^2) = \prod_{i=1}^n f(\xi_i; a, \sigma^2)$  досягає свого найбільшого значення. Чи є  $a^*$  та  $(\sigma^2)^*$  оцінками і, якщо так, чи будуть вони незміщеними оцінками параметрів  $a$  та  $\sigma^2$  відповідно? Конзистентними оцінками цих параметрів?

**1.10.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta, \sigma^2) = \begin{cases} \frac{1}{2\sqrt{3}\sigma}, & \text{якщо } x \in [\theta - \sqrt{3}\sigma, \theta + \sqrt{3}\sigma]; \\ 0, & \text{якщо } x \notin [\theta - \sqrt{3}\sigma, \theta + \sqrt{3}\sigma]. \end{cases}$$

(рівномірний на відрізку  $[\theta - \sqrt{3}\sigma, \theta + \sqrt{3}\sigma]$  розподіл). Позначимо через  $\theta^*$ ,  $(\sigma^2)^*$  розв'язок системи рівнянь

$$\begin{cases} m_1(\theta^*; (\sigma^2)^*) = \frac{1}{n} \sum_{i=1}^n \xi_i; \\ m_2(\theta^*; (\sigma^2)^*) = \frac{1}{n} \sum_{i=1}^n \xi_i^2, \end{cases}$$

де  $m_k(\theta; \sigma^2) = \int_{\mathbb{R}} x^k f(x; \theta, \sigma^2) dx$ ,  $k = \overline{1, 2}$ .

Чи є  $\theta^*$  та  $(\sigma^2)^*$  оцінками і, якщо так, чи є вони незміщеними оцінками параметрів  $\theta$  та  $\sigma^2$ ? Конзистентними оцінками цих параметрів?

**1.11.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу

$$P(k; \theta) = \frac{1}{1+\theta} \left( \frac{\theta}{1+\theta} \right)^k, k = 0, 1, \dots; \theta > 0.$$

Чи є

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$$

ефективною оцінкою параметра  $\theta$ ?

**Зауваження.** Якщо визначити  $\frac{1}{1+\theta} = p$ , то розподіл

$$P(k; \theta) = \frac{1}{1+\theta} \left( \frac{\theta}{1+\theta} \right)^k, k = 0, 1, \dots$$

набуває добре відомого вигляду геометричного розподілу

$$P(k; p) = p(1 - p)^k, k = 0, 1, \dots$$

**1.12.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з нормального розподілу зі щільністю

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

Позначимо через  $a^*, (\sigma^2)^*$  розв'язок системи рівнянь

$$\begin{cases} m_1(a^*, (\sigma^2)^*) = \frac{1}{n} \sum_{i=1}^n \xi_i; \\ m_2(a^*, (\sigma^2)^*) = \frac{1}{n} \sum_{i=1}^n \xi_i^2, \end{cases}$$

де  $m_k(a; \sigma^2) = \int_{\mathbb{R}} x^k f(x; a, \sigma^2) dx, k = \overline{1, 2}$ .

Чи є  $a^*$  та  $(\sigma^2)^*$  оцінками? І якщо так, то чи є  $a^*$  і  $(\sigma^2)^*$  незміщеними оцінками параметрів  $a$  та  $\sigma^2$ ? Конзистентними оцінками цих параметрів?

**1.13.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta) = \begin{cases} \frac{x^{m-1} e^{-\frac{x}{\theta}}}{\theta^m (m-1)!}, & \text{якщо } x > 0; (\theta > 0) \\ 0, & \text{якщо } x \leq 0 \end{cases}$$

(розподіл Ерланга),  $m$  – відоме. Чи є

$$\theta^* = \frac{1}{mn} \sum_{i=1}^n \xi_i$$

ефективною оцінкою параметра  $\theta$ ?

**1.14.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & \text{якщо } x > 0; \\ 0, & \text{якщо } x \leq 0, \end{cases} \quad (\theta > 0)$$

(показниковий розподіл з параметром  $\frac{1}{\theta}$ ). Чи є

$$\theta^* = \frac{1}{n} \sum_{i=1}^n \xi_i$$

ефективною оцінкою параметра  $\theta$ ?

**1.15.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x-m_0}{\theta}}, & \text{якщо } x > m_0; \\ 0, & \text{якщо } x \leq m_0, \end{cases} \quad (\theta > 0, m_0 \text{ – відоме}).$$

Чи є  $\theta^* = \bar{\xi} - m_0$  ефективною оцінкою параметра  $\theta$ ?

**1.16.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \alpha) = \begin{cases} \frac{1}{\alpha} e^{-\frac{x}{\alpha}}, & \text{якщо } x > 0; \\ 0, & \text{якщо } x \leq 0. \end{cases}$$

Довести, що

$$a_1^* = \frac{1}{n} \sum_{i=1}^n \xi_i$$

є незміщеною і конзистентною оцінкою параметра  $\alpha$ . Чи є

$$a_2^* = \frac{1}{2} (\xi_{n-1} + \xi_n)$$

незміщеною оцінкою параметра  $\alpha$ ? Конзистентною оцінкою цього параметра?

**1.17.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з біноміального розподілу

$$P(k; \theta) = C_N^k \theta^k (1 - \theta)^{N-k}, \quad k = \overline{0, N},$$

де  $N$  – відоме натуральне число. Чи є

$$\theta^* = \frac{1}{nN} \sum_{i=1}^n \xi_i$$

незміщеною та конзистентною оцінкою параметра  $\theta$ ?

**1.18.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta, b) = \begin{cases} \frac{1}{\theta} e^{-\frac{x-b}{\theta}}, & \text{якщо } x > b; \\ 0, & \text{якщо } x \leq b. \end{cases}$$

Розглядаються оцінки:

$$\theta_1^* = \frac{1}{n} \sum_{i=1}^n \xi_i; \theta_2^* = \min\{\xi_i\}; \theta_3^* = \theta_1^* + \theta_2^*.$$

Чи є серед  $\theta_1^*$ ,  $\theta_2^*$ ,  $\theta_3^*$  незміщені, конзистентні оцінки параметрів  $\theta$  та  $b$ ? Можливо, серед них є незміщені та конзистентні оцінки інших параметрів? Яких саме?

**1.19.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta) = \begin{cases} \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}, & \text{якщо } x > 0; (\theta > 0). \\ 0, & \text{якщо } x \leq 0, \end{cases}$$

(розподіл Релея). Чи є

$$\theta^* = \frac{1}{2n} \sum_{i=1}^n \xi_i^2$$

ефективною оцінкою параметра  $\theta$ ?

**1.20.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; b) = \begin{cases} \frac{1}{b-a_0}, & \text{якщо } x \in [a_0, b]; \\ 0, & \text{якщо } x \notin [a_0, b]. \end{cases} \quad a_0 \text{ – відоме.}$$

Розглядаються оцінки:

$$\theta_1^* = \min\{\xi_i\}; \theta_2^* = \max\{\xi_i\}; \theta_3^* = \frac{1}{n} \sum_{i=1}^n \xi_i;$$

$$\theta_4^* = \max\{\xi_i\} + \frac{1}{n-1}(\max\{\xi_i\} - a_0).$$

Серед  $\theta_i^*$ ,  $i = \overline{1,4}$ , визначити незміщені, конзистентні оцінки параметра  $b$ . Можливо, серед них є незміщені й конзистентні оцінки інших параметрів? Яких саме?

**1.21.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з біноміального розподілу з параметрами  $m$  і  $p$ :

$$P(k; p) = C_m^k p^k (1-p)^{m-k}, \quad k = \overline{0, m}.$$

Чи є

$$p^* = \frac{1}{mn} \sum_{i=1}^n \xi_i$$

ефективною оцінкою параметра  $p$ ? Конзистентною оцінкою цього параметра?

**1.22.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta) = \begin{cases} \frac{1}{2}, & \text{якщо } x \in [\theta - 1; \theta + 1]; \\ 0, & \text{якщо } x \notin [\theta - 1; \theta + 1]. \end{cases}$$

Розглядаються оцінки:

$$\theta_1^* = \min\{\xi_i\}; \quad \theta_2^* = \max\{\xi_i\}; \quad \theta_3^* = \frac{1}{n} \sum_{i=1}^n \xi_i; \quad \theta_4^* = \frac{\xi_n + \xi_{n+1}}{2};$$

$$\theta_5^* = \frac{1}{2}(\max\{\xi_i\} + \min\{\xi_i\}); \quad \theta_6^* = \min\{\xi_i\} - \frac{1}{n-1}(\max\{\xi_i\} - \min\{\xi_i\});$$

$$\theta_7^* = \max\{\xi_i\} + \frac{1}{n-1} (\max\{\xi_i\} - \min\{\xi_i\}).$$

Визначити серед  $\theta_i^*$ ,  $i = \overline{1,7}$ , незміщені, конзистентні оцінки параметра  $\theta$ . Можливо, серед них є незміщені, конзистентні оцінки інших параметрів? Яких саме?

**1.23.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta) = \begin{cases} \frac{x^{v-1} e^{-\frac{x}{\theta}}}{\theta^v \Gamma(v)}, & \text{якщо } x > 0; \\ 0, & \text{якщо } x \leq 0. \end{cases}$$

(гамма-розподіл);  $v$  – відомий параметр. Чи є

$$\theta^* = \frac{1}{nv} \sum_{i=1}^n \xi_i$$

ефективною оцінкою параметра  $\theta$ ?

**1.24.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta, h) = \begin{cases} \frac{1}{h}, & \text{якщо } x \in [\theta - h; \theta + h]; \\ 0, & \text{якщо } x \notin [\theta - h; \theta + h]. \end{cases}$$

Розглядаються оцінки:

$$\theta_1^* = \min\{\xi_i\}; \theta_2^* = \max\{\xi_i\}; \theta_3^* = \frac{1}{n} \sum_{i=1}^n \xi_i; \theta_4^* = \frac{\xi_n + \xi_{n-1}}{2};$$

$$\theta_5^* = \max\{\xi_i\} - \min\{\xi_i\}; \theta_6^* = \frac{1}{2} (\max\{\xi_i\} + \min\{\xi_i\});$$

$$\theta_7^* = \min\{\xi_i\} - \frac{1}{n-1} (\max\{\xi_i\} - \min\{\xi_i\});$$

$$\theta_8^* = \max\{\xi_i\} + \frac{1}{n-1} (\max\{\xi_i\} - \min\{\xi_i\}).$$

Визначити серед  $\theta_i^*$ ,  $i = \overline{1,8}$ , незміщені, конзистентні оцінки параметрів  $a$  та  $b$ . Можливо, серед них є незміщені, конзистентні оцінки інших параметрів? Яких саме?

**1.25.** Відомо, що  $\theta^*$  є незміщеною точковою оцінкою невідомого параметра  $\theta$  розподілу генеральної сукупності. Чи буде точкова оцінка  $(\theta^*)^2$  незміщеною оцінкою  $\theta^2$ ?

**1.26.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з гауссівського розподілу з параметрами  $a$  ( $a$  – відомий параметр) та  $\sigma^2$ . Чи буде статистика

$\sigma^* = \frac{1}{n} \sqrt{\frac{\pi}{2}} \sum_{k=1}^n |\xi_k - a|$  незміщеною та конзистентною оцінкою параметра  $\sigma$ ?

**1.27.** Нехай  $\xi_1$  та  $\xi_2$  два спостереження гауссівської генеральної сукупності з параметрами  $0$  та  $\sigma^2$ . Чи буде статистика  $\sigma^* = \frac{\sqrt{\pi}}{2} |\xi_1 + \xi_2|$  незміщеною оцінкою параметра  $\sigma$ ?

**1.28.** Нехай  $\xi_1$  та  $\xi_2$  два спостереження гауссівської генеральної сукупності з параметрами  $a$  та  $\sigma^2$ . Чи буде статистика  $\sigma^* = \frac{\sqrt{\pi}}{2} |\xi_1 - \xi_2|$  незміщеною оцінкою параметра  $\sigma$ ?

## Глава 2. Методи побудови точкових оцінок

### 2.1. Емпіричні оцінки

*Емпірична функція розподілу.* Як відомо, закон розподілу будь-якої випадкової величини  $\xi$  можна задавати функцією розподілу  $F_\xi(x)$ . Якщо мати вибірку значень цієї випадкової величини, то можна побудувати деяке наближення функції розподілу.

*Емпіричною функцією розподілу* називається функція розподілу дискретної випадкової величини, що набуває значення варіант  $x_k$  з ймовірністю  $\omega_k$ ,  $k = \overline{1, m}$ , тобто:

$$F_n^*(x) = \begin{cases} 0, & \text{при } x \leq x_1; \\ \omega_k^{\text{нак}}, & \text{при } x_k < x \leq x_{k+1}; \\ 1, & \text{при } x > x_m, \end{cases}$$

де  $\omega_k^{\text{нак}} = \sum_{r=1}^k \omega_r$ , тобто накопичена частість.

Очевидно, що  $F_n^*(x) = 0$  при  $x \leq x_1$  і  $F_n^*(x) = 1$  при  $x > x_m$ . На проміжку  $(x_1, x_m]$   $F_n^*(x)$  є неспадною кусково-сталогою функцією.

Аналогічно визначається емпірична функція розподілу для інтервального варіаційного ряду. Оскільки, ми маємо лише значення емпіричної функції  $F_n^*(x)$  на кінцях інтервалів, то для графічного її зображення доцільно до визначити  $F_n^*(x)$ , з'єднавши точки графіка, які відповідають кінцям інтервалів, відрізками прямої. В результаті отримана ламана співпадає з кумулятою.

Із теореми Бернуллі випливає, що емпірична функція розподілу  $F_n^*(x)$  при  $n \rightarrow \infty$  прямує за ймовірністю до функції розподілу  $F_\xi(x)$  в кожній точці неперервності  $F_\xi$ . Крім того, візьмемо до уваги *теорему*

*Глівенка - Кантеллі*, згідно якої,  $F_n^*(x) \rightarrow F_\xi(x)$  при  $n \rightarrow \infty$  із ймовірністю 1 рівномірно відносно  $x$ .

***Емпіричні (вибіркові) характеристики генеральної сукупності.***

Нехай задано вибірковий вектор  $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$  з генеральної сукупності випадкової величини  $\xi$ . Функція  $h = h(\vec{\xi}) = h(\xi_1, \xi_2, \dots, \xi_n)$  від вибіркового вектора  $\vec{\xi}$  називається ***статистикою***. Найбільш розповсюджені статистики – це емпіричні (вибіркові) моменти.

***Вибірковим моментом порядку  $k$***  називається величина

$$E_k^* = m_k^* = \int_{\mathbb{R}} x^k d F_n^*(x) = \frac{1}{n} \sum_{i=1}^n \xi_i^k.$$

***Вибірковим центральним моментом порядку  $k$***  називається величина

$$\tilde{E}_k^* = \mu_k^* = \int_{\mathbb{R}} (x - m_1^*)^k d F_n^*(x) = \frac{1}{n} \sum_{i=1}^n (\xi_i - m_1^*)^k.$$

Вибірковий момент першого порядку найчастіше називають ***вибірковим (емпіричним) середнім*** і позначають

$$m_1^* = \bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i.$$

Вибіркове середнє є незміщеною оцінкою математичного сподівання генеральної сукупності.

Реалізацію вибіркового середнього будемо позначати  $\bar{x}$ . Якщо  $x_i$  є варіантами дискретного варіаційного ряду,  $n_i$  – відповідні їм частоти ( $i = \overline{1, m}$ ), то вибіркове середнє обчислюється так:

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

У випадку інтервального варіаційного ряду за значення  $x_i$  беруть середини інтервалів варіаційного ряду. Очевидно, що у цьому випадку  $\bar{x} = \sum_{i=1}^m x_i \omega_i$ , де  $\omega_i$  – відповідні частоти варіантів або інтервалів ( $i = \overline{1, m}$ ).

Вибірковий центральний момент другого порядку називають **вибірковою (емпіричною) дисперсією** та позначають

$$\mu_2^* = D^* \xi = \frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})^2.$$

Зауважимо, що вибіркову дисперсію можна записати так  $D^* \xi = \overline{\xi^2} - (\bar{\xi})^2 = \frac{1}{n} \sum_{i=1}^n \xi_i^2 - (\bar{\xi})^2$ . Вибіркова дисперсія є зміщеною оцінкою дисперсії. В якості незміщеної оцінки дисперсії потрібно брати **виправлену вибіркову дисперсію**  $\tilde{D} \xi = \frac{n}{n-1} D^* \xi$ . Крім вибіркової дисперсії розглядають також **вибіркове середнє квадратичне** або **стандартне відхилення**  $\sigma_\xi^* = \sqrt{D^* \xi}$ . Розглянуті вище статистики називають **аналітичними**. Аналітичними оцінками також є емпіричний коефіцієнт асиметрії та емпіричний ексцес. У математичній статистиці застосовують також **порядкові** статистики. До таких статистик належать емпірична мода, емпірична медіана,

**Емпіричною (вибірковою) модою**  $Mo^* \xi$  називається варіанта, якій відповідає найбільша частість. Особливість моди в тому, що вона не змінюється при зміні крайніх членів ряду, тобто має певну стійкість до варіації.

Отже, для дискретного варіаційного ряду  $(Mo^*\xi)_{зн} = x_k$ , якщо  $n_k = \max_i n_i$ , а для інтервального варіаційного ряду

$$(Mo^*\xi)_{зн} = y_i + h_i \frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}},$$

де  $y_i$  – початок інтервалу з найбільшою частотою,  $n_i$  – частота  $i$ -го інтервалу.

**Емпіричною (вибірковою) медіаною  $Me^*\xi$**  називають середню за розташуванням варіанту дискретного варіаційного ряду, якщо кількість варіант – непарна, і середнім арифметичним двох середніх за розташуванням варіант, якщо кількість варіант – парна. Тобто реалізацію медіани обчислюють так:

$$(Me^*\xi)_{зн} = \begin{cases} x_{(\frac{m+1}{2})}, & \text{при } m = 2n - 1; \\ \frac{x_{(\frac{m}{2})} + x_{(\frac{m}{2}+1)}}{2}, & \text{при } m = 2n. \end{cases}$$

Для інтервальних варіаційних рядів

$$(Me^*\xi)_{зн} = y_i + h_i \frac{\frac{n}{2} - \sum_{k=1}^{i-1} n_k}{n_i},$$

де  $y_i$  – початок медіанного інтервалу, тобто такого, якому відповідає перша з частот, що перевищує половину всіх спостережень,  $h_i$  – довжина інтервалу,  $n_i$  – частота медіанного інтервалу.

**Емпіричним (вибірковим) коефіцієнтом асиметрії** називають статистику  $As^*\xi = \frac{\mu_3^*}{(\mu_2^*)^{3/2}}$ , а **емпіричним (вибірковим) ексцесом** –

$$Ex^*\xi = \frac{\mu_4^*}{(\mu_2^*)^2} - 3.$$

### Приклади розв'язання задач

**Приклад 2.1.1.** Обчислити вибіркве середнє, виправлену вибіркву дисперсію, вибіркву моду, вибіркву медіани для вибірки

інтервал	[-2; 0)	[0; 4)	[4; 6)	[6; 10]
$n_i$	5	10	20	15

**Розв'язання.** Значення вибіркового середнього будемо обчислювати за формулою

$$\bar{x} = \frac{1}{n} \sum_{i=1}^4 x_i^* n_i, \text{ де } x_i^* - \text{середина інтервалу } [x_{i-1}; x_i), (i = \overline{1, 4}).$$

$$\bar{x} = \frac{1}{50} (-5 + 20 + 100 + 120) = 4,7.$$

Для обчислення виправленої вибіркової дисперсії спочатку обчислимо

$$\frac{1}{n} \sum_{i=1}^4 (x_i^*)^2 n_i = \frac{1}{50} ((-1)^2 \cdot 5 + 2^2 \cdot 10 + 5^2 \cdot 20 + 8^2 \cdot 15) = 30,1,$$

тоді значення вибіркової дисперсії дорівнює

$$D^*\xi_{\text{зн}} = 30,1 - (4,7)^2 = 8,01.$$

Значення виправленої вибіркової дисперсії буде наступним

$$\tilde{D}\xi_{\text{зн}} = \frac{50}{49} \cdot 8,01 = 8,17.$$

Згідно наведених вище формул, обчислюємо значення вибіркової моди та медіани

$$(Me^* \xi)_{\text{зн}} = 4 + 2 \cdot \frac{25-15}{20} = 5.$$

$$(Mo^* \xi)_{\text{зн}} = 4 + 2 \cdot \frac{20-10}{40-10-15} = 5,33.$$

## 2.2. Метод моментів

Ідея *методу моментів*, який був запропонований К.Пірсоном, полягає в тому, що певну кількість емпіричних моментів, які обчислюються за вибіркою, прирівнюють до відповідних теоретичних моментів, які обчислені за законом генеральної сукупності.

Тобто емпіричні початкові моменти  $m_k^* = \frac{1}{n} \sum_{i=1}^n \xi_i^k$  прирівнюють до теоретичних початкових моментів  $m_k = E\xi^k = \int_{\mathbb{R}} x^k dF(x)$ . Із отриманого рівняння (або системи рівнянь) визначають оцінку невідомого параметра (або параметрів)  $\theta$ . Слід зазначити, що деякі теоретичні моменти можуть не залежати від невідомих параметрів. У такому випадку до виписаних рівнянь додають наступні, щоб при цьому кількість рівнянь дорівнювала кількості невідомих параметрів.

### *Приклади розв'язання задач*

**Приклад 2.2.1.** Нехай випадкова величина  $\xi$  має гамма розподіл зі щільністю

$$f(x, \lambda, \alpha) = \begin{cases} \frac{\lambda^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}, & \text{при } x \geq 0; \\ 0, & \text{при } x < 0, \end{cases}$$

де  $\lambda$  і  $\alpha$  — два невідомих параметра.

Знайти за допомогою методу моментів оцінки невідомих параметрів  $\lambda$  і  $\alpha$ .

**Розв'язання.** Помітимо, що цьому розподілу підпорядковується час  $\xi$  до відмови системи із  $\alpha = m$  ( $m \in \mathbb{N}$ ) однотипних елементів, якщо кожен із  $m - 1$  елементів вмикається в роботу після відмови попереднього, і час до відмови  $\xi_i, i = \overline{1, m}$ , будь-якого елемента має експоненційний розподіл

$$f(x; \lambda) = \begin{cases} \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, & \text{при } x \geq 0; \\ 0, & \text{при } x < 0. \end{cases}$$

У даному випадку, використовуючи визначення гамма-функції

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt,$$

а також рекурентне співвідношення  $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$ , отримаємо наступний вираз для першого, другого початкових моментів і дисперсії:

$$m_1 = \int_0^\infty \frac{\lambda^\alpha x^\alpha}{\Gamma(\alpha)} e^{-\lambda x} dx = \frac{\Gamma(\alpha+1)}{\lambda\Gamma(\alpha)} = \frac{\alpha}{\lambda},$$

$$m_2 = \int_0^\infty \frac{\lambda^\alpha x^{\alpha+1}}{\Gamma(\alpha)} e^{-\lambda x} dx = \frac{\Gamma(\alpha+2)}{\lambda^2\Gamma(\alpha)} = \frac{\alpha(\alpha+1)}{\lambda^2},$$

$$D\xi = E(\xi^2) - (E\xi)^2 = m_2 - m_1^2 = \frac{\alpha}{\lambda^2}.$$

Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка обсягу  $n$  із генеральної сукупності  $\xi$ . Знаходимо значення вибіркового середнього  $\bar{x}$  та вибіркової дисперсії  $D^*\xi_{\text{зн}}$ . Прирівнюючи моменти  $m_1 = M\xi$  і  $m_2 = D\xi$  до існуючих моментів вибірки, отримуємо систему рівнянь

$$\begin{cases} \frac{\alpha}{\lambda} = \bar{x}; \\ \frac{\alpha}{\lambda^2} = D^*\xi_{\text{зн}}, \end{cases}$$

звідки знаходимо значення оцінок  $\lambda_{\text{зн}}^* = \frac{\bar{x}}{D^*\xi_{\text{зн}}}$ ,  $\alpha_{\text{зн}}^* = \left(\frac{\bar{x}}{\sigma_{\text{зн}}^*}\right)^2$ .

Отже, оцінками невідомих параметрів будуть статистики

$$\lambda^*(\vec{\xi}) = \frac{\bar{\xi}}{D^*\xi}, \quad \alpha^*(\vec{\xi}) = \left(\frac{\bar{\xi}}{\sigma^*}\right)^2.$$

**Приклад 2.2.2.** Методом моментів знайдемо оцінку параметра  $\theta = p$  в біноміальній моделі, де  $p$  є ймовірність «успіху» в будь-якому із  $n$  незалежних спостережень, що повторюються, а випадкова величина  $k(\vec{\xi})$  – число «успіхів».

**Розв'язання.** Випадковою вибіркою  $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$  в даному випадку є  $n$  дискретних випадкових величин  $\xi_i$ , кожна з яких приймає значення 1 з ймовірністю  $p$  і 0 з ймовірністю  $1 - p$  (біномний розподіл). При цьому  $k(\vec{\xi}) = \xi_1 + \dots + \xi_n$ , а математичне сподівання  $Ek(\vec{\xi}) = np$ .

Якщо в результаті  $n$  незалежних спостережень ми отримали вибіркоче значення  $k(\vec{\xi}) = k$ , то рівняння, яке потрібно скласти згідно методу моментів, має вигляд  $np = k$ .

Отримаємо  $p^* = \frac{k}{n}$ . Отже, точковою оцінкою параметра  $p$  є відносна частота.

### 2.3. Метод максимальної правдоподібності

Нехай  $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$  – вибірка з генеральної сукупності із законом розподілу  $F_{\xi}(x, \theta)$ , що залежить від параметра  $\theta$ . Параметр  $\theta \in \Theta$  невідомий і його треба оцінити за вибіркою  $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ . Найважливішим, як у теоретичному так і у практичному планах, методом побудови оцінок є **метод максимальної правдоподібності**, запропонований Р. Фішером.

Метод максимальної правдоподібності полягає в тому, що за оцінку невідомого параметра  $\theta$  вибирається така точка  $\theta^*$ , в якій функція правдоподібності  $L(\vec{\xi}, \theta)$  досягає свого максимального значення

$$L(\vec{\xi}, \theta^*) = \max_{\theta \in \Theta} L(\vec{\xi}, \theta).$$

Таким чином, за методом максимальної правдоподібності в якості оцінки невідомого параметра розподілу  $\theta$  набувають таке значення  $\theta^*$ , яке максимізує функцію правдоподібності. Такий підхід у знаходженні оцінок впливає зі змісту функції правдоподібності, яка при кожному фіксованому значенні параметра  $\theta$  є мірою правдоподібності отримання реалізації вибірки  $\vec{x} = (x_1, x_2, \dots, x_n)$ . Отже, за оцінку параметра  $\theta$

береться таке значення, для якого спостереження  $\vec{x} = (x_1, x_2, \dots, x_n)$  є найбільш імовірними.

Зазначимо також, що функції  $L(\vec{\xi}, \theta)$  та  $\ln L(\vec{\xi}, \theta)$  досягають максимуму в одній точці. Але ж відшукувати максимальне значення  $\ln L(\vec{\xi}, \theta)$  зручніше ніж  $L(\vec{\xi}, \theta)$ , тому найчастіше шукають максимум саме  $\ln L(\vec{\xi}, \theta)$ .

Рівняння вигляду  $\frac{\partial \ln L(\vec{\xi}, \theta)}{\partial \theta} = 0$  називається *рівнянням максимальної правдоподібності*, а функція  $\ln L(\vec{\xi}, \theta)$  – *логарифмічною функцією правдоподібності*.

Якщо потрібно знайти оцінки декількох параметрів розподілу  $\theta_1, \theta_2, \dots, \theta_s$ , то розв'язують  $s$  рівнянь правдоподібності

$$\frac{\partial \ln L(\vec{\xi}, \theta_1, \theta_2, \dots, \theta_s)}{\partial \theta_i} = 0, i = \overline{1, s}.$$

### **Приклади розв'язання задач**

**Приклад 2.3.1.** Нехай величина  $\xi$ , що спостерігається в експерименті – час роботи приладу до відмови – має експоненційний розподіл зі щільністю

$$f(x, \lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{при } x \geq 0; \\ 0, & \text{при } x < 0, \end{cases}$$

де  $\lambda$  – невідомий параметр.

Застосовуючи метод максимальної правдоподібності, знайдемо точкову оцінку для параметра  $\lambda$ .

**Розв'язання.** Нехай  $\vec{x} = (x_1, \dots, x_n)$  – будь-яка реалізація випадкової вибірки  $\vec{\xi}$  із генеральної сукупності  $\xi$ .

У випадку, що розглядається

$$L(x_1, \dots, x_n; \lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n \exp(-\lambda \sum_{i=1}^n x_i),$$
$$\ln L(x_1, \dots, x_n; \lambda) = n \ln \lambda - \lambda \sum_{i=1}^n x_i.$$

Отже, рівняння правдоподібності має вигляд

$$\frac{\partial \ln L(\vec{x}; \lambda)}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0,$$

звідки витікає, що

$$\lambda_{\text{ЗН}}^* = \left( \frac{1}{n} \sum_{i=1}^n x_i \right)^{-1}.$$

Отже точковою оцінкою невідомого параметра  $\lambda \in \lambda^*(\vec{\xi}) = \frac{1}{\bar{\xi}}$ .

Якщо врахувати, що  $E\xi = \frac{1}{\lambda}$ , а найкращою оцінкою  $E\xi$  є вибіркоче середнє  $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$ , то отримана відповідь є цілком природною.

**Приклад 2.3.2.** Для загальної нормальної моделі  $N(\theta_1, \theta_2^2)$  методом максимальної правдоподібності знайдемо оцінку вектора параметрів  $\vec{\theta} = (\theta_1, \theta_2)$ .

**Розв'язання.** У цьому випадку функція правдоподібності

$$L(\vec{x}; \theta_1, \theta_2) = \frac{1}{(\theta_2 \sqrt{2\pi})^n} \exp\left(-\frac{1}{2\theta_2^2} \sum_{i=1}^n (x_i - \theta_1)^2\right)$$

і, як наслідок,

$$\ln L(\vec{x}; \theta_1, \theta_2) = -n \ln \sqrt{2\pi} - n \ln \theta_2 - \frac{1}{2\theta_2^2} \sum_{i=1}^n (x_i - \theta_1)^2.$$

Оскільки кількість невідомих параметрів  $r = 2$ , система рівнянь правдоподібності буде складатись із двох рівнянь

$$\begin{cases} \frac{\partial}{\partial \theta_1} \ln L(\vec{x}; \theta_1, \theta_2) = \frac{1}{\theta_2^2} \sum_{i=1}^n (x_i - \theta_1) = 0, \\ \frac{\partial}{\partial \theta_2} \ln L(\vec{x}; \theta_1, \theta_2) = -\frac{n}{\theta_2} + \frac{1}{\theta_2^3} \sum_{i=1}^n (x_i - \theta_1)^2 = 0. \end{cases}$$

Розв'язуючи систему, отримаємо

$$(\theta_1^*)_{\text{ЗН}} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (\theta_2^*)_{\text{ЗН}}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Отже, оцінками максимальної правдоподібності для математичного сподівання  $E\xi = \theta_1$  і дисперсії  $D\xi = \theta_2^2$  випадкової величини, що розподілена за нормальним законом, є відповідно вибіркове середнє

$$\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$$

і вибіркова дисперсія

$$D^* \xi = \frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})^2.$$

## Задачі для самостійного розв'язання

**2.1.** Обчислити незміщені оцінки математичного сподівання та дисперсії, вибірку медіану та моду для вибірки

а)

$x_i$	12	14	16	18	20	22
$n_i$	5	15	50	16	10	4

б)

інтервал	[2; 4)	[4; 6)	[6; 10)	[10; 16)	[16; 20]
$n_i$	2	8	35	40	15

**2.2.** Побудувати емпіричну функцію розподілу, полігон частот. Обчислити незміщені оцінки математичного сподівання та дисперсії, вибірку медіану та моду для вибірки

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

**2.3.** Як зміняться значення вибіркового середнього, значення вибіркової моди та вибіркової медіани, якщо кожний член вибірки:

а) збільшити (зменшити) на число  $d$ ;

б) збільшити (зменшити) в  $k$  разів?

**2.4.** Як зміниться значення вибіркової дисперсії, якщо кожний член вибірки:

а) збільшити (зменшити) на число  $d$ ;

б) збільшити (зменшити) в  $k$  разів?

**2.5.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з геометричного розподілу з параметром  $p$ .

$$P(k; p) = p(1 - p)^k, \quad k = 0, 1, \dots$$

Оцінити параметр  $p$  методом моментів.

**2.6.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; a, b) = \frac{1}{2a} e^{-\frac{1}{a}|x-b|}, \quad a > 0.$$

Знайти оцінки параметрів  $a$  і  $b$  методом максимальної правдоподібності.

**2.7.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta) = \frac{1}{2\theta} e^{-\frac{1}{\theta}|x|}.$$

Знайти оцінку дисперсії методом моментів і з'ясувати, чи є вона незміщеною та конзистентною оцінкою дисперсії.

**2.8.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з рівномірного на відріжку  $[\theta - \sigma\sqrt{3}, \theta + \sigma\sqrt{3}]$  розподілу,  $\sigma > 0$ . Знайти оцінки  $\theta^*$  та  $(\sigma^2)^*$  параметрів  $\theta$  і  $\sigma^2$  методом моментів. З'ясувати, чи є оцінки  $\theta^*$  та  $(\sigma^2)^*$  незміщеними і конзистентними оцінками відповідних параметрів.

**2.9.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta) = \begin{cases} \frac{x}{\theta} e^{-\frac{x^2}{2\theta}}, & \text{якщо } x > 0; (\theta > 0) \\ 0, & \text{якщо } x \leq 0, \end{cases}$$

(розподіл Релея).

Знайти оцінку параметра  $\theta$  методом максимальної правдоподібності. З'ясувати, чи є оцінка максимальної правдоподібності параметра  $\theta$  його незміщеною оцінкою? Конзистентною оцінкою? Ефективною оцінкою?

**2.10.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу Ерланга з параметрами  $m$  та  $\lambda$ . Оцінити ці параметри методом моментів. Розподіл Ерланга з параметрами  $m$  і  $\lambda$  визначається як розподіл суми  $m$  незалежних випадкових величин, кожна з яких має показників розподіл з параметром  $\lambda$ . Щільність розподілу Ерланга з параметрами  $m$  та  $\lambda$  має вигляд

$$f(x; \lambda, m) = \begin{cases} \frac{\lambda^m}{(m-1)!} x^{m-1} e^{-\lambda x}, & \text{якщо } x > 0; \\ 0, & \text{якщо } x \leq 0. \end{cases}$$

**2.11.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; a, b) = \begin{cases} \frac{1}{a} e^{-\frac{x-b}{a}}, & \text{якщо } x > b; \\ 0, & \text{якщо } x \leq b \end{cases}$$

(зміщений експоненційний розподіл).

Знайти оцінки параметрів  $a$  і  $b$  методом максимальної правдоподібності. З'ясувати, чи є оцінки максимальної правдоподібності параметрів  $a$  та  $b$  незміщеними оцінками? Конзистентними оцінками?

**2.12.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \nu, \alpha) = \begin{cases} \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x}, & \text{якщо } x > 0; \\ 0, & \text{якщо } x \leq 0. \end{cases} \quad (\nu \geq 0, \alpha > 0)$$

(гамма розподіл).

Знайти оцінку параметрів  $\nu$  та  $\alpha$  методом моментів.

**2.13.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу

$$P(k; \theta) = \frac{1}{1+\theta} \left( \frac{\theta}{1+\theta} \right)^k, \quad k = 0, 1, \dots; \theta > 0$$

(розподіл Паскаля).

Знайти оцінки параметра  $\theta$  методом максимальної правдоподібності. З'ясувати, чи є оцінка максимальної правдоподібності незміщеною, конзистентною та ефективною оцінкою параметра  $\theta$ .

**2.14.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

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

(нормальний розподіл з параметрами  $a; \sigma^2$ ).

Знайти оцінку параметрів  $a$  та  $\sigma^2$  методом моментів.

**2.15.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з біноміального розподілу

$$P(k; m) = C_m^k p^k (1 - p)^{m-k}, \quad k = \overline{0, m}.$$

Оцінити параметри  $m$  та  $p$  методом моментів.

**2.16.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу зі щільністю

$$f(x; \theta, \lambda) = \begin{cases} \frac{\theta \lambda^\theta}{x^{\theta+1}}, & \text{якщо } x > \lambda; \\ 0, & \text{якщо } x \leq \lambda. \end{cases} \quad (\theta > 2, \lambda > 0)$$

(розподіл Парето).

Знайти оцінку параметрів  $\theta$  та  $\lambda$  методом моментів.

**2.17.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу

$$P(k; \theta) = C_{r-1+k}^{r-1} \left(\frac{1}{1+\theta}\right)^r \left(\frac{\theta}{1+\theta}\right)^k, \quad k = 0, 1, \dots; \theta > 0, r \in \mathbb{N} \text{ – відоме.}$$

Знайти оцінку  $\theta^*$  параметра  $\theta$  методом моментів. З'ясувати, чи є  $\theta^*$  незміщеною, конзистентною та ефективною оцінкою параметра  $\theta$ .

**2.18.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з біноміального розподілу з параметрами  $m$  і  $p$ :

$$P(k; m) = C_m^k p^k (1 - p)^{m-k}, \quad k = \overline{0, m}, \quad m \text{ – відомий параметр.}$$

Оцінити параметр  $p$  методом максимальної правдоподібності. З'ясувати, чи є оцінка максимальної правдоподібності параметра  $p$  незміщеною, конзистентною та ефективною оцінкою.

**2.19.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка з розподілу Пуассона

$$P(k; \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

Знайти оцінку параметра  $\lambda$  методом максимальної правдоподібності. З'ясувати, чи є оцінка максимальної правдоподібності параметра  $\lambda$  його незміщеною, ефективною, та конзистентною оцінкою.

**2.20.** Із генеральної сукупності, що розподілена за біномним законом, отримана вибірка обсягу  $n$  (дивись приклад 2.2.2). Знайдіть методом максимальної правдоподібності оцінку невідомого параметра  $p$  і покажіть, що ця оцінка буде незміщеною, конзистентною і ефективною.

**2.21.** Знайдіть методом максимальної правдоподібності за вибіркою обсягупточкову оцінку геометричного розподілу

$$P\{\xi = k\} = p(1 - p)^{k-1}, \quad k \in \mathbb{N},$$

де  $\xi$  – число дослідів до появи події;  $p$  – ймовірність появи події в одному досліді.

**2.22.** Знайдіть методом максимальної правдоподібності за вибіркою обсягу  $n$  точкову оцінку параметра  $\beta$  гамма-розподілу ( $\alpha$  відоме) зі щільністю

$$f(x, \alpha, \beta) = \begin{cases} \frac{x^\alpha e^{-\frac{x}{\beta}}}{\beta^{\alpha+1} \Gamma(\alpha+1)}, & \text{якщо } x > 0; (\alpha > 1, \beta > 0). \\ 0, & \text{якщо } x \leq 0 \end{cases}$$

**2.23.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка із генеральної сукупності  $\xi$ , що розподілена за законом зі щільністю

$$f(x, \alpha) = \begin{cases} \frac{\alpha^k x^{k-1} e^{-\alpha x}}{\Gamma(k)}, & \text{якщо } x > 0; \\ 0, & \text{якщо } x \leq 0. \end{cases}$$

де  $\alpha > 0$  – невідомий параметр,  $k \in \mathbb{N}$  відоме. Знайдіть за допомогою методу максимальної правдоподібності оцінку параметра  $\alpha$ .

**2.24.** Нехай  $\xi_1, \xi_2, \dots, \xi_n$  – вибірка розподілу зі щільністю

$$f(x, \theta) = \begin{cases} \frac{e^{-|x|}}{2(1-e^{-\theta})}, & \text{якщо } |x| \leq \theta; (\theta > 0). \\ 0, & \text{якщо } |x| > \theta \end{cases}$$

Знайдіть оцінку максимальної правдоподібності для параметра  $\theta$ .

### Глава 3. Побудова довірчих інтервалів для параметрів закону розподілу

Точкова оцінка  $\theta_n^*$  невідомого параметра розподілу  $\theta$ , як правило, з точним значенням параметра не співпадає. Незміщені оцінки співпадають з параметром  $\theta$  лише в середньому, конзистентні оцінки прямують до параметра  $\theta$  зі збільшенням обсягу вибірки, а ефективні оцінки мають найменшу степінь випадкових відхилень від  $\theta$ . Отже, завжди існує похибка  $|\theta_n^* - \theta| > 0$ . Величина похибки при цьому невідома, хоча потрібно знати, до яких помилок призведе заміна параметра  $\theta$  на його точкову оцінку  $\theta_n^*$ . Щоб отримати уявлення про точність і надійність точкової оцінки  $\theta_n^*$  для параметра  $\theta$ , використовують інтервальне оцінювання параметрів розподілу.

*Інтервальною оцінкою* параметра  $\theta$  називають інтервал  $(\theta_1^*, \theta_2^*)$ , який з наперед заданою ймовірністю  $\gamma = 1 - \varepsilon$  накриває невідоме значення параметра  $\theta$ , тобто

$$P\{\theta_1^* < \theta < \theta_2^*\} \geq \gamma.$$

Одразу звернемо увагу на те, що межі інтервалу  $(\theta_1^*, \theta_2^*)$  знаходяться за вибіркою, тому є випадковими величинами на відміну від параметра  $\theta$  – величини не випадкової. Інтервал  $(\theta_1^*, \theta_2^*)$ , який з наперед заданою ймовірністю  $\gamma = 1 - \varepsilon$  накриває невідоме значення параметра, називається *довірчим інтервалом*, величина  $\gamma$  – *довірчою ймовірністю* або *надійністю* оцінки, а величина  $\varepsilon$  – *рівнем значущості* оцінки.

Величина довірчого інтервалу суттєво залежить від обсягу вибірки  $n$  (зменшується зі зростанням  $n$ ) та від значення довірчої ймовірності  $\gamma$  (збільшується при зростанні  $\gamma$  до одиниці). Отже, надійність прийнято вибирати рівною 0,9, 0,95, 0,99 тощо. Тоді подія, яка полягає в тому, що довірчий інтервал  $(\theta_1^*, \theta_2^*)$  накриє точне значення параметра  $\theta$ , є практично вірогідною.

Слід зауважити, що інколи довірчий інтервал вибирається симетричним відносно параметра  $\theta$ , тобто  $(\theta^* - \delta, \theta^* + \delta)$ . У таких випадках величину  $\delta$  називають точністю оцінки  $\theta^*$ .

Найбільш просто довірчі інтервали можна знайти для математичного сподівання, дисперсії у випадках, коли генеральна сукупність має нормальний розподіл. Розглянемо саме такі випадки.

Нехай  $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$  є вибірка з генеральної сукупності випадкової величини  $\xi$ , розподіленої за нормальним законом з параметрами  $a$  та  $\sigma^2$ . Поставимо задачу побудови довірчого інтервалу для параметра  $a$  з наперед заданою довірчою ймовірністю  $\gamma$ . Розділимо задачу на два випадки: 1) дисперсія генеральної сукупності – відома величина; 2) дисперсія генеральної сукупності – невідома.

Отже, нехай відома дисперсія  $\sigma^2$  випадкової величини  $\xi$ . За вибіркою знайдемо вибіркоче середнє  $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$ . Вибіркове середнє нормально розподіленої генеральної сукупності є нормально розподіленою випадковою величиною з математичним сподіванням  $a$  та дисперсією  $\frac{\sigma^2}{n}$ . Задамо ймовірність  $\gamma$  та знайдемо величину  $\delta$ , для якої ймовірність події

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^n \xi_i - a \right| < \delta \right\} = \gamma.$$

Із формул попадання нормально розподіленої випадкової величини в заданий інтервал

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^n \xi_i - a \right| < \delta \right\} = 2\Phi \left( \frac{\delta\sqrt{n}}{\sigma} \right) = \gamma.$$

За таблицею значень функції Лапласа знаходимо аргумент  $t_{\gamma/2} = \Phi^{-1} \left( \frac{\gamma}{2} \right)$ . Звідси величина  $\delta = \frac{t_{\gamma/2}\sigma}{\sqrt{n}}$ . Це значення задає точність шуканого довірчого інтервалу, який в цьому випадку набуває вигляд  $(\bar{\xi} - \delta, \bar{\xi} + \delta)$ .

Розв'яжемо задачу побудови довірчого інтервалу для математичного сподівання нормально розподіленої випадкової величини  $\xi$  у випадку невідомої дисперсії. За вибіркою знайдемо вибіркове середнє  $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$  та виправлену вибіркову дисперсію  $\tilde{D}\xi = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2$ . Задамо довірчу ймовірність  $\gamma$ . Довірчий інтервал для математичного сподівання випадкової величини  $\xi$  також будемо шукати у вигляді  $P \{ |\bar{\xi} - E\xi| < \delta \} = \gamma$ . Зауважимо, що статистика  $\frac{(\bar{\xi} - E\xi)\sqrt{n}}{\sqrt{\tilde{D}\xi}}$ , має розподіл Стюдента з  $n - 1$  ступенями вільності. Отже, за таблицею значень розподілу Стюдента, при заданій надійності  $\gamma$  і числа  $k = n - 1$ , визначимо величину  $t_\gamma$  таку, що  $P \left\{ \left| \frac{(\bar{\xi} - E\xi)\sqrt{n}}{\sqrt{\tilde{D}\xi}} \right| < t_\gamma \right\}$ . Із цієї рівності випливає, що

$$P \left\{ \bar{\xi} - \frac{t_\gamma \sqrt{\tilde{D}\xi}}{\sqrt{n}} < E\xi < \bar{\xi} + \frac{t_\gamma \sqrt{\tilde{D}\xi}}{\sqrt{n}} \right\}.$$

Отже, з імовірністю  $\gamma$  можна стверджувати, що вибіркоче середнє дає значення невідомого математичного сподівання з точністю  $\delta = \frac{t_\gamma \sqrt{\tilde{D}\xi}}{\sqrt{n}}$ , і довірчий інтервал має вигляд

$$\left( \bar{\xi} - \frac{t_\gamma \sqrt{\tilde{D}\xi}}{\sqrt{n}}, \bar{\xi} + \frac{t_\gamma \sqrt{\tilde{D}\xi}}{\sqrt{n}} \right).$$

Поставимо задачу побудови довірчого інтервалу для параметра  $\sigma^2$  з наперед заданою довірчою ймовірністю  $\gamma$ . Як відомо, точковою оцінкою дисперсії є вибіркоче дисперсія  $D^*\xi = \frac{1}{n} \sum_{i=1}^n (\xi_i - E\xi)^2$ , якщо математичне сподівання відоме, та виправлена вибіркоче дисперсія  $\tilde{D}\xi = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2$ , якщо математичне сподівання випадкової величини  $\xi$  – невідоме. У випадку відомого математичного сподівання, статистика  $\chi_n^2 = \frac{nD^*\xi}{\sigma^2}$  має розподіл хі-квадрат з  $n$  ступенями вільності. Аналогічно, статистика  $\chi_{n-1}^2 = \frac{(n-1)\tilde{D}\xi}{\sigma^2}$  має розподіл хі-квадрат з  $n-1$  ступенями вільності, якщо математичне сподівання невідоме.

Розглянемо випадок відомого математичного сподівання. При побудові довірчого інтервалу для дисперсії знайдемо  $\sigma^2 = \frac{nD^*\xi}{\chi_n^2}$ . Крива розподілу величини  $\chi_n^2$  має вигляд (рис. 3.1), тобто не симетрична відносно початку координат.

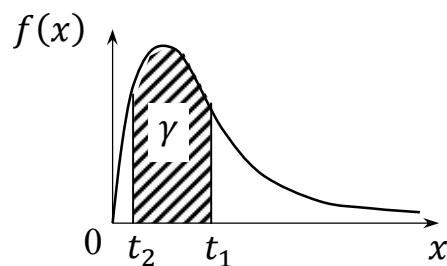


Рисунок 3.1.

Виберемо інтервал  $I_\gamma$ , в який попадає випадкова величина  $\chi_n^2$  з імовірністю  $\gamma$  так, щоб імовірності виходу величини  $\chi_n^2$  за границі інтервалу вправо та вліво були однакові й дорівнювали  $\frac{1-\gamma}{2}$ :

$$P\{\chi_n^2 < t_2\} = \frac{1-\gamma}{2} \Leftrightarrow P\{\chi_n^2 > t_2\} = \frac{1+\gamma}{2} \text{ та}$$

$$P\{\chi_n^2 > t_1\} = \frac{1-\gamma}{2} \Leftrightarrow P\{\chi_n^2 < t_1\} = \frac{1+\gamma}{2}.$$

Для знаходження чисел  $t_1$  та  $t_2$  використаємо таблицю значень розподілу хі-квадрат, взявши  $n$  ступенів вільності. Нерівності  $\chi_n^2 < t_1$  та  $\chi_n^2 > t_2$  еквівалентні нерівностям  $\sigma^2 > \frac{nD^*\xi}{t_1}$  та  $\sigma^2 < \frac{nD^*\xi}{t_2}$ . Отже, шуканий довірчий інтервал для дисперсії нормально розподіленої випадкової величини у випадку відомого математичного сподівання має вигляд

$$\left( \frac{nD^*\xi}{t_1}, \frac{nD^*\xi}{t_2} \right).$$

Якщо математичне сподівання невідоме, то величини  $t_1$  та  $t_2$  знаходимо також із таблиці розподілу хі-квадрат по значенням  $\frac{1-\gamma}{2}$  та  $\frac{1+\gamma}{2}$ , взявши  $n - 1$  ступінь вільності. Довірчий інтервал при цьому набуває вигляд

$$\left( \frac{(n-1)\tilde{D}\xi}{t_1}, \frac{(n-1)\tilde{D}\xi}{t_2} \right).$$

**Наближені інтервальні оцінки.** Якщо розподіл генеральної сукупності невідомий, а точкова оцінка параметра є асимптотично нормальною, то для побудови наближеного довірчого інтервалу для

параметра користуються центральною граничною теоремою (див. приклад 3.1.6).

### *Приклади розв'язання задач*

**Приклад 3.1.1.** За допомогою вольтметра, точність якого характеризується середнім квадратичним відхиленням  $0,2$  В, проведено  $10$  вимірів напруги бортової батареї. Знайти довірчий інтервал для істинного значення напруги батареї з довірчою ймовірністю  $\gamma = 0,95$ , якщо середнє арифметичне результатів спостережень  $\bar{x} = 50,2$  В. Контрольована ознака має нормальний закон розподілу.

**Розв'язання.** Для знаходження довірчого інтервалу

$$\left( \bar{x} - t_{\gamma/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + t_{\gamma/2} \frac{\sigma}{\sqrt{n}} \right),$$

де значення  $t_{\gamma/2}$  визначається за допомогою таблиці функції Лапласа  $\Phi(t_{\gamma/2}) = \frac{\gamma}{2}$ . За цією таблицею  $t_{\gamma/2} = t_{0,475} = 1,96$ . Оскільки  $t_{\gamma/2} \frac{\sigma}{\sqrt{n}} = 1,96 \frac{0,2}{\sqrt{10}} \approx 0,1$ , довірчий інтервал має вигляд  $(50,2 - 0,1; 50,2 + 0,1)$ , або  $(50,1; 50,3)$ .

**Приклад 3.1.2.** З великої партії електроламп було відібрано випадковим чином  $400$  штук для середньої тривалості горіння. Вибіркове середнє тривалості горіння ламп виявилася рівним  $1220$  г. Знайти з довірчою ймовірністю  $\gamma = 0,9973$  довірчий інтервал для середньої

тривалості горіння електроламп по всій партії, якщо середнє квадратичне відхилення тривалості горіння – 35 г.

**Розв’язання.** Незалежно від закону розподілу генеральної сукупності  $\xi$  (тривалість горіння електролампи) статистика  $\frac{\bar{\xi}-\mu}{\sigma}\sqrt{n}$ , де  $\bar{\xi} = \frac{1}{n}\sum_{i=1}^n \xi_i$  має асимптотично нормальний розподіл з параметрами  $(0, 1)$ , що впливає з центральної граничної теореми.

Оскільки об’єм вибірки великий ( $n = 400$ ), то межі довірчого інтервалу знаходимо так само, як і в прикладі 3.1.1.

Для  $\gamma = 0,9973$  за допомогою таблиці функції Лапласа знаходимо.  $t_{\gamma/2} = u_{0,49865} = 3$ . В силу співвідношення  $t_{\gamma/2} \frac{\sigma}{\sqrt{n}} = \frac{3 \cdot 35}{\sqrt{400}} \approx 5,25$ , довірчий інтервал має вигляд  $(1220 - 5,25; 1220 + 5,25)$ , або  $(1214,75; 1225,25)$ .

**Приклад 3.1.3.** У результаті запусків 10 ракет отримані (в умовних одиницях) значення бокових відхилень точок попадання від точок прицілу (табл. 3.1).

Таблиця 3.1

Номер ракети	1	2	3	4	5	6	7	8	9	10
Відхилення	1,0	0,2	1,0	-0,1	-0,5	5,0	-1,0	3,0	0,5	1,0

Вважаючи, що випадкова величина  $\xi$  (випадкове бокове відхилення точок попадання від точок прицілу) має нормальний розподіл, побудувати довірчий інтервал для її математичного сподівання з довірчою ймовірністю  $\gamma = 0,99$ .

**Розв'язання.** Для знаходження довірчого інтервалу скористуємося статистикою  $\frac{(\bar{\xi}-\mu)\sqrt{n-1}}{\sigma^*(\bar{\xi})}$ , яка має розподіл Стюдента з  $n - 1$  ступенями вільності. Вибіркове середнє має значення

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{10} (1 + 0,2 + 1 - 0,1 - 0,5 + 5 - 1 + 3 + 0,5 + 1) = 1,01,$$

а вибіркова дисперсія – значення

$$\begin{aligned} D^* \xi_{\text{зн}} = S^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \\ &= \frac{1}{10} \left( (-0,01)^2 + 0,99^2 + (-0,01)^2 + (-1,11)^2 + (-1,51)^2 + 3,99^2 + \right. \\ &\quad \left. + (-2,01)^2 + 1,99^2 + (-0,51)^2 + (-0,01)^2 \right) = \\ &= 2,8349. \end{aligned}$$

Значення вибіркового середнього квадратичного відхилення  $\sigma_{\text{зн}}^* = S = \sqrt{2,8349} \approx 1,68$ . За таблицею квантилів розподілу Стюдента для  $n - 1 = 9$  знаходимо квантиль

$$t_{1-\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n-1}} = t_{0,005}(9) \frac{1,68}{\sqrt{9}} = 3,25 \cdot \frac{1,68}{3} \approx 1,82,$$

отримаємо довірчий інтервал  $(1,01 - 1,82; 1,01 + 1,82)$ , або  $(-0,81; 2,83)$ .

**Приклад 3.1.4.** Із партії однотипних високоомних опорів відібрано 10 штук. У кожного із них виміряне відхилення опорів від номінального значення (табл. 3.2).

Таблиця 3.2

Номер виробу	1	2	3	4	5	6	7	8	9	10
Відхилення	1	3	-2	2	4	2	5	3	-2	4

Вважаючи, що контрольована ознака має нормальний має нормальний закон розподіл, довірчий інтервал для дисперсії з довірчою ймовірністю  $\gamma = 0,98$ .

**Розв'язання.** Знаходимо значення вибіркового середнього

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1+3-2+2+4+2+5+3-2+4}{10} = 2$$

і значення виправленої вибіркової дисперсії

$$\tilde{D}_{\xi_{\text{зн}}} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1+1+16+4+9+1+1+16+4}{9} \approx 5,88.$$

Щоб побудувати довірчий інтервал для дисперсії, скористаємося статистикою

$$\frac{(n-1)\tilde{D}_{\xi}}{\sigma^2} = \frac{nD^*\xi}{\sigma^2},$$

яка має розподіл хі-квадрат з  $n - 1$  ступенями вільності. За таблицею квантилів даного розподілу знаходимо квантили  $\chi_{\frac{\alpha}{2}, n-1}^2$  і  $\chi_{1-\frac{\alpha}{2}, n-1}^2$  рівнів  $\frac{\alpha}{2}$  і  $1 - \frac{\alpha}{2}$ . У даному випадку

$$\alpha = 1 - \gamma = 1 - 0,98 = 0,02$$

і розподіл має дев'ять ступенів вільності. Отже,

$$\chi_{\frac{\alpha}{2}, 9}^2 = \chi_{0,01; 9}^2 = 2,09; \quad \chi_{1-\frac{\alpha}{2}, 9}^2 = \chi_{0,99; 9}^2 = 21,67.$$

Для меж довірчого інтервалу отримаємо

$$\frac{(n-1)\tilde{D}\xi_{\text{зн}}}{\chi_{1-\frac{\alpha}{2}, n-1}^2} = \frac{5,88 \cdot 9}{21,67} = 2,45; \quad \frac{(n-1)\tilde{D}\xi_{\text{зн}}}{\chi_{\frac{\alpha}{2}, n-1}^2} = \frac{5,88 \cdot 9}{2,09} = 25,36.$$

Звідси знаходимо довірчий інтервал для дисперсії з довірчим коефіцієнтом 0,98: (2,45; 25,36).

**Приклад 3.1.5.** Знайдемо довірчий інтервал для ймовірності попадання снаряду в ціль з довірчою ймовірністю  $\gamma = 0,9$ , якщо після 220 пострілів у ціль попало 75 снарядів.

**Розв'язання.** Використовуючи таблицю функції Лапласа, знаходимо значення  $t_{\gamma/2} = t_{0,45} = 1,645$ . Межі довірчого інтервалу мають вигляд

$$\bar{p}_1 = \frac{m}{n} - \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sqrt{\frac{m}{n} \left(1 - \frac{m}{n}\right)} = \frac{75}{220} - \frac{1,645}{\sqrt{220}} \sqrt{\frac{75 \cdot 145}{220^2}} \approx 0,288,$$

$$\bar{p}_2 = \frac{m}{n} + \frac{u_{1-\frac{\alpha}{2}}}{\sqrt{n}} \sqrt{\frac{m}{n} \left(1 - \frac{m}{n}\right)} = \frac{75}{220} + \frac{1,645}{\sqrt{220}} \sqrt{\frac{75 \cdot 145}{220^2}} \approx 0,393.$$

Отже, довірчий інтервал для ймовірності попадання снаряду в ціль має вигляд: (0,288; 0,393).

**Приклад 3.1.6.** Знайти довірчий інтервал для параметра  $a$  закону Пуассона.

**Розв'язання.** Розглянемо вибірку  $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$  із закону Пуассона з параметром  $a$ . За точкову оцінку параметра  $a$  доречно взяти вибіркове середнє  $\bar{\xi}$ , оскільки математичне сподівання  $E\xi = a$ . При великому обсязі вибірки вибіркове середнє має наближено нормальний закон розподілу  $N\left(a, \sqrt{\frac{a}{n}}\right)$ . Таким чином, статистика  $t = \frac{(\bar{\xi}-a)\sqrt{n}}{\sqrt{a}}$  має наближено стандартний нормальний розподіл  $N(0, 1)$ .

Користуючись таблицею значень функції Лапласа, для заданої довірчої ймовірності  $\gamma$  знайдемо таке число  $t_{\gamma/2}$ , при якому

$$P\left\{\left|\frac{(\bar{\xi}-a)\sqrt{n}}{\sqrt{a}}\right| < t_{\gamma/2}\right\} = \gamma.$$

Розв'язуючи цю нерівність відносно величини  $a$ , отримаємо  $a \in (a_1, a_2)$ , де  $a_{1,2} = \bar{\xi} + \frac{t_{\gamma/2}^2}{2n} \mp \frac{t_{\gamma/2}}{2n} \sqrt{t_{\gamma/2}^2 + 4n\bar{\xi}}$ .

### Задачі для самостійного розв'язання

**3.1.** Оцінка величини, що вимірюється визначається формулою вибірковим середнім  $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$ . Результати окремих вимірів не містять систематичної похибки і підпорядковуються нормальному закону розподілу з середнім квадратичним відхиленням  $\sigma = 2,1$ . Визначте інтервальні оцінки з довірчою ймовірністю 0,9 для значення величини, що

вимірюється при різних об'ємах випадкової вибірки  $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ :  
а)  $n = 5$ , б)  $n = 10$ , в)  $n = 25$ .

**3.2.** Середня квадратична похибка висотоміра  $\sigma = 15$  м. Скільки потрібно мати таких приладів на літаку, щоб з імовірністю 0,99 похибка виміру середньої висоти  $\bar{\xi}$  була менша ніж 30 м? При цьому випадкові помилки розподілені за нормальним законом, а систематичні помилки відсутні.

**3.3.** На основі 100 дослідів було визначено, що в середньому для виробництва деталі потрібно  $\bar{t} = 5,5$  с, а  $\bar{\sigma}_t = 1,7$  с. Зробивши припущення, що час для виготовлення деталі розподілений за нормальним законом, визначте довірчий інтервал для математичного сподівання виробництва деталі з довірчою ймовірністю 0,85.

**3.4.** За результатами вимірів 100 резисторів, випадково відібраних з великої партії однотипних виробів, отримана оцінка опору  $\bar{x} = 10$  кОм. Знайдіть:

а) ймовірність того, що для резисторів всієї партії значення опору лежать у межах  $(10 \pm 0,1)$  кОм (середнє квадратичне відхилення виміру відоме:  $\sigma = 1$  кОм);

б) кількість вимірів, при яких з ймовірністю 0,95 можна стверджувати, що для всієї партії резисторів значення опорів лежать у межах  $(10 \pm 0,1)$  кОм.

**3.5.** Провели 5 незалежних рівноточних вимірів для визначення заряду електрона; отримали наступні результати (в абсолютних

електростатичних одиницях):  $4,781 \cdot 10^{-10}; 4,792 \cdot 10^{-10}; 4,795 \cdot 10^{-10}; 4,779 \cdot 10^{-10}; 4,769 \cdot 10^{-10}$ . Визначте значення оцінки величини заряду електрона і знайти довірчий інтервал при надійності 99%, вважаючи, що помилки розподілені за нормальним законом і виміри не мають систематичних помилок.

**3.6.** На контрольних дослідах 16 освітлювальних ламп були визначені значення оцінок математичного сподівання і середнього квадратичного відхилення їх терміну служби, які виявились рівними  $\bar{x} = 3000$  г. і  $\bar{\sigma} = 20$  г. відповідно. Вважаючи, що контрольована ознака (термін служби лампи) має нормальний закон розподілу, визначте:

а) довірчий інтервал для математичного сподівання  $m$  при довірчій ймовірності 0,9;

б) ймовірність, з якої можна стверджувати, що абсолютна величина похибки визначення  $m$  не перевищить 10 г.

**3.7.** Провели 20 вимірювань бази довжиною  $L$ . За результатом досліду отримані значення оцінок величини, що вимірюється і середнього квадратичного відхилення:  $\bar{x} = 10400$  м і  $\bar{\sigma}_x = 85$  м. Похибки виміру підпорядковуються нормальному закону розподілу. Знайдіть ймовірність того, що інтервал з випадковими межами  $(0,999 \bar{\xi}, 1,001 \bar{\xi})$  накрис невідомий параметр  $L$ .

**3.8.** Із партії валів відібрали  $n_1 = 9$  штук. Значення вибіркового середнього діаметру вала  $\bar{x}_1 = 30$  мм, вибіркової дисперсії  $S_1^2 = 9$  мм<sup>2</sup>. Потім зробили повторний експеримент, відібравши  $n_2 = 16$  шт. і отримали значення вибірових оцінок  $\bar{x}_2 = 29$  мм,  $S_2^2 = 4,5$  мм<sup>2</sup>. Використовуючи

об'єднані вибірккові оцінки, знайдіть 99%-ий довірчий інтервал для середнього.

**3.9.** За результатами 10 вимірів ємності конденсатора приладом, який немає систематичної похибки, отримали наступні відхилення від номінального значення (пФ):

5,4; -13,9; -11; 7,2; -15,6; 29,2; 1,4; -0,3; 6,6; -9,9.

Знайдіть 90%-ий довірчий інтервал для дисперсії, припускаючи, що генеральна сукупність має нормальний розподіл.

**3.10.** По 15 незалежним рівноточним вимірам були розраховані значення оцінок математичного сподівання і середнього квадратичного відхилення максимальної швидкості літака  $\bar{v} = 424,7$  м/с і  $\bar{\sigma}_v = 7,7$  м/с. Вважаючи, що генеральна сукупність має нормальний розподіл, визначте: а) довірчий інтервал для середнього квадратичного відхилення при довірчій ймовірності 0,9; б) ймовірність того, що абсолютна величина випадкової похибки при визначенні  $\sigma_v$  по 15 вимірам не перевищить 2 м/с.

**3.11.** Відомо, що вимірювальний пристрій не має систематичних похибок, а випадкові похибки виміру підпорядковуються нормальному закону розподілу. Скільки потрібно провести вимірів для визначення оцінки середнього квадратичного відхилення приладу, щоб з довірчою ймовірністю 0,7 абсолютна величина похибки визначення цієї величини була не більша 20% від  $\bar{\sigma}(\xi)$ ?

**3.12.** При перевірці 100 деталей із великої партії виявлено 10 бракованих. Знайдіть 95%-ий довірчий інтервал для кількості бракованих деталей у всій партії.

**3.13.** Із великої партії з  $N$  транзисторів одного типу було випадково вибрано і перевірено 100 шт. Коефіцієнт підсилення 36 транзисторів виявився менше 10. Знайдіть 95%-ий довірчий інтервал для кількості таких транзисторів у всій партії.

**3.14.** З автоматичної лінії, що виробляє підшипники, було відібрано 100 шт., причому 10 виявилися бракованими. Знайдіть: а) 90%-ий довірчий інтервал для ймовірності того, що навмання вибраний підшипник виявиться бракованим; б) кількість підшипників, які потрібно перевірити, щоб з ймовірністю 0,9973 можна було стверджувати, що доля браку відрізняється від частоти не більше ніж на 0,05.

**3.15.** У 10000 сеансах гри з автоматом виграш з'явиться 4000 раз. Знайдіть: а) 95% довірчий інтервал для ймовірності виграшу; б) кількість сеансів гри, які потрібно провести, щоб з ймовірністю 0,99 можна було стверджувати, що ймовірність  $p$  виграшу відрізняється від її частоти не більше ніж на 1%.

**3.16.** Для експоненційного розподілу зі «зсувом», що має щільність

$$f(x; \theta) = \begin{cases} e^{\theta-x}, & \text{якщо } x \geq \theta; \\ 0, & \text{якщо } x < \theta, \end{cases}$$

за вибіркою обсягу  $n$  побудуйте інтервальну оцінку параметра  $\theta$  з довірчою ймовірністю  $\gamma$ .

**3.17.** Проведено п'ять спостережень гауссівської випадкової величини:

$$-2,5; 3,4; -2,0; 1,0; 2,1.$$

Знайти оцінку математичного сподівання та побудувати для неї 90%-ий довірчий інтервал.

**3.18.** На контрольних дослідженнях 16 ламп були визначені значення найкращих точкових оцінок терміну їх роботи:  $\bar{x} = 3000$  год.,  $\bar{\sigma} = 20$  год. Вважаючи, що термін роботи кожної лампи має нормальний розподіл, визначити: а) довірчий інтервал для математичного сподівання, який з ймовірністю 0,9 накриває значення математичного сподівання; б) з якою надійністю можна стверджувати, що абсолютне значення похибки визначення математичного сподівання не перевищить 10 год.?

**3.19.** Точкова оцінка дисперсії, що отримана при обробці результатів 100 незалежних спостережень випадкової величини  $\xi$ , дорівнює 20. Знайти: а) довірчий інтервал для дисперсії випадкової величини  $\xi$ , що відповідає довірчій ймовірності 0,9; б) надійність оцінки дисперсії, що обчислена з точністю  $\varepsilon = 3$  ( $\varepsilon$  - половина довжини довірчого інтервалу); в) мінімально необхідну кількість досліджень для того, щоб оцінити дисперсію з точністю  $\varepsilon = 3,92$  при надійності  $\gamma = 0,95$ .

## Глава 4. Критерій Пірсона ( $\chi^2$ )

Інформація, що отримана при обробці вибірки з деякої генеральної сукупності, може бути використана для отримання висновків про всю генеральну сукупність. Подібні висновки називають *статистичними*. Завдяки їх імовірнісному характеру завжди можна знайти ймовірність того, що прийняте рішення буде помилковим, тобто оцінити ступінь ризику того чи іншого прийнятого рішення.

*Статистичною гіпотезою  $H$*  називається припущення про вигляд розподілу генеральної сукупності, яке перевіряється за вибіркою. Часто розподіл генеральної сукупності відомий і за вибіркою треба перевірити припущення щодо значень параметрів цього розподілу. Такі гіпотези називають *параметричними*. Деякі з них розглядаються в наступній главі. Гіпотеза, що перевіряється, називається *нульовою (основною)* гіпотезою та позначається  $H_0$ . Решта гіпотез називається *альтернативними (конкуруючими)* відносно нульової гіпотези та позначаються  $H_1, H_2, \dots$ . Правило, за яким приймається чи відхиляється гіпотеза на основі вибірки, називається *критерієм* для перевірки гіпотези. Якщо перевіряється гіпотеза про належність розподілу генеральної сукупності до якогось класу розподілів (нормального, рівномірного, Пуассона, тощо), то зазначене правило називається *критерієм згоди*.

Оскільки висновок щодо прийняття або відхилення гіпотези формується за вибіркою, то вибране рішення може бути помилковим. Розрізняють такі типи помилок. Помилка, яка полягає в тому, що правильна гіпотеза  $H_0$ , згідно з вибраним критерієм, відхиляється, називається *помилкою першого роду*. *Помилка другого роду* відбувається

тоді, коли справджується альтернативна гіпотеза, але приймається основна гіпотеза  $H_0$ .

Ймовірність  $\alpha$  допустити помилку першого роду називається рівнем значущості, ймовірність  $1 - \beta$  не допустити помилку другого роду називається потужністю критерію. У цій главі всюди, якщо не вказано інакше, за рівень значущості  $\alpha$  береться 0,05.

#### 4.1. Перевірка гіпотези про вид розподілу генеральної сукупності за критерієм $\chi^2$

Перейдемо до найважливішої задачі математичної статистики: встановлення теоретичного закону розподілу генеральної сукупності за певною реалізацією вибірки.

Припущення про вид закону генеральної сукупності варто висувати тільки після первинної обробки статистичних даних. Параметри розподілу, як правило, невідомі, тому їх замінюємо на найкращі точкові оцінки. Очевидно, що між теоретичними та емпіричними розподілами існують розходження. Важливо зрозуміти, чи пояснюються ці розходження тільки випадковими обставинами (наприклад, обмеженість кількості спостережень), чи суттєвими (наприклад, теоретичний закон підібрано невдало).

На практиці частіше всього використовується *критерій Пірсона* ( $\chi^2$ ), який ґрунтується на *теоремі Пірсона*:

Якщо складна гіпотеза  $H_0$  про закон розподілу генеральної сукупності справджується, то статистика критерію  $\eta = \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i}$  прямує в слабкому розумінні до розподілу  $\chi^2$  (Пірсона) з  $r - s - 1$  ступенями вільності.

*Пояснення.* В наведеній теоремі  $n$  є обсяг вибірки. Згідно висунутій гіпотезі множину значень, що може приймати генеральна сукупність  $\xi$  розбиваємо на  $r$  підмножин  $X = \bigcup_{i=1}^r X_i$ , які попарно не перетинаються. Обчислюємо (за висуненою гіпотезою)  $p_i = P(\xi \in X_i / H_0)$ ,  $\sum_{i=1}^r p_i = 1$ . Число  $s$  означає кількість параметрів гіпотетичного закону, що були оцінені,  $n_i$  кількість значень реалізації вибірки, що попали в  $X_i$ ,  $i = \overline{1, r}$ .

Алгоритм використання теореми Пірсона на практиці складається з таких кроків.

1. Після висування гіпотези множину значень генеральної сукупності  $\xi$  розбиваємо на  $r$  підмножин  $X_i$ , які попарно не перетинаються.

2. Згідно висунутій гіпотези, підраховуємо  $p_i$ . Перевіряємо умову  $\sum_{i=1}^r p_i = 1$ . Якщо  $r \geq 20$ , то потрібно виконання умови  $np_i \geq 5$ ; якщо ж  $r < 20$ , то потрібно виконання умови  $np_i \geq 10$ . Це пов'язано з тим, що розподіл статистики критерію  $\eta$  замінюється на розподіл  $\chi^2$ . Якщо відповідні умови не виконуються, то сусідні підмножини  $X_i$  об'єднують. Це впливає на значення  $r$ .

3. Підраховуємо значення статистики критерію  $\eta_{zn}$  з коректованою кількістю ступенів вільності і порівнюємо з критичним значенням  $t_{\alpha, k}$ , яке беремо з таблиці розподілу Пірсона

4. Критична область в критерії Пірсона є правосторонньою. Отже, якщо  $\eta_{zn} < t_{\alpha, k}$ , то на рівні значущості  $\alpha$ , дані не суперечать висунутій гіпотезі  $H_0$ ; якщо ж  $\eta_{zn} > t_{\alpha, k}$ , то на рівні значущості  $\alpha$ , дані суперечать висунутій гіпотезі  $H_0$ .

Зазначимо що на практиці окрім критерію  $\chi^2$  використовують критерій Колмогорова, в якому за міру розходження між теоретичними та емпіричними розподілами розглядають максимальне значення абсолютної величини різниці між емпіричної та гіпотетичної функціями розподілу.

Але критерієм Колмогорова можна користуватись лише тоді, коли гіпотетичний розподіл повністю визначений та неперервний.

### *Приклади розв'язання задач*

У цій главі всюди, якщо не вказано інакше, за рівень значущості  $\alpha$  береться 0,05.

#### **Приклад 4.1.1. Покази механічних годинників.**

У таблиці 4.1 наведено покази 500 навмання вибраних годинників, виставлених у вітринах годинникових магазинів. Чи узгоджується з наведеними даними гіпотеза про те, що покази годинників рівномірно розподілені на відріжку  $[0;12)$ ?

*Таблиця 4.1*

$i$	0	1	2	3	4	5	
$n_i$	41	34	54	39	49	45	
$i$	6	7	8	9	10	11	Разом
$n_i$	41	33	37	41	47	39	500

Тут  $i$  – номер проміжку від  $i$ -ї години до  $(i + 1)$ -ї,  $i = 0, 1, \dots, 11$ ;  $n_i$  – кількість годинників, покази яких належать  $i$ -му відріжку.

**Розв'язання.** Покази 500 годинників можна розглядати як реалізацію вибірки обсягом 500 з деякого неперервного на відріжку  $[0;12)$  розподілу  $F$ . Щодо розподілу  $F$  випадкової величини  $\xi$  (показів

годинників) висувається гіпотеза  $H_0$ :  $F$  є рівномірним розподілом на відріжку  $[0; 12)$ , тобто щільність  $f(x)$  розподілу  $F$  має вигляд

$$f(x) = \begin{cases} \frac{1}{12}, & \text{якщо } x \in [0,12); \\ 0, & \text{якщо } x \notin [0,12). \end{cases}$$

Для перевірки гіпотези  $H_0$  скористаємося критерієм  $\chi^2$ .

Поділимо множину можливих значень  $X = [0; 12)$  випадкової величини  $\xi$  на перетині підмножини  $X_i = [i; i + 1), i = 0, 1, \dots, 11$ . Імовірність потрапляння  $\xi$  в кожен з цих множин

$$p_i = P\{\xi \in [i, i + 1)\} = \int_i^{i+1} \frac{1}{12} dx = \frac{1}{12},$$

причому,

$$np_i = 500 \frac{1}{12} = 41,7 > 10; \quad i = 0, 1, \dots, 11.$$

Отже, користуватися критерієм  $\chi^2$  можна.

Обчислимо значення відхилення між емпіричним розподілом та гіпотетичним:

$$\begin{aligned} \eta_{\text{ЗН}} &= \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i} = \\ &= \frac{1}{41,7} ((41 - 41,7)^2 + \dots + (39 - 41,7)^2) = 9,99; \\ \eta_{\text{ЗН}} &= 9,99 < 19,7 = t_{0,05; 11} = t_{\alpha; (r-1)}. \end{aligned}$$

Тому, згідно з критерієм  $\chi^2$ , гіпотеза про рівномірний на відріжку  $[0; 12)$  розподіл випадкової величини  $\xi$  – показів механічних годинників у

вітринах магазинів – не відхиляється. Дані таблиці характерні для вибірки обсягом 500 з рівномірного на відріжку  $[0; 12)$  розподілу. Припущення про те, що покази годинників розподілені рівномірно на цьому відріжку, не суперечить спостереженням.

**Приклад 4.1.2.** Серед 2020 дводітних сімей 527 мають двох хлопчиків і 476 – двоє дівчаток, у решти (1017 сімей) діти різної статі.

Чи можна вважати, що кількість хлопчиків у сім'ї, яка має двох дітей, є біномно розподіленою випадковою величиною?

**Розв'язання.** Розглянемо випадкову величину  $\xi$ , яка для кожної дводітної сім'ї, набуває значення  $i$ , якщо в сім'ї  $i$  хлопчиків,  $i = \overline{0, 1, 2}$ . Щодо розподілу випадкової величини  $\xi$  висувається гіпотеза

$$H_0: P\{\xi = i\} = C_2^i p^i (1 - p)^{2-i}, \quad i = \overline{0, 1, 2}.$$

Потрібно перевірити цю гіпотезу.

Параметр  $p$  гіпотетичного розподілу невідомий. Його необхідно оцінити за вибіркою. Оцінкою максимальної правдоподібності параметра  $p$  біномного розподілу

$$C_m^k p^k (1 - p)^{m-k}, \quad k = \overline{0, m}.$$

за вибіркою  $\xi_1, \xi_2, \dots, \xi_n \in p^* = \frac{1}{mn} \sum_{k=1}^n \xi_k$ .

Зокрема, у розглядуваному прикладі значення оцінки  $p^*$  параметра  $p$

$$p^* = \frac{0 \cdot 476 + 1 \cdot 1017 + 2 \cdot 527}{2 \cdot 2020} = 0,513.$$

Тому гіпотетичний розподіл можна записати ще й так:

$$G = \begin{pmatrix} 0 & 1 & 2 \\ 0,237 & 0,500 & 0,263 \end{pmatrix}.$$

За поділ вибіркового простору  $X = \{0; 1; 2\}$  визначаємо множини  $X_0 = \{0\}$ ,  $X_1 = \{1\}$ ,  $X_2 = \{2\}$ . Для цих множин  $np_i \geq 10$ ,  $i = \overline{0,1,2}$ , де  $\bar{p}_i$ ,  $i = \overline{0,1,2}$ , – оцінки гіпотетичних ймовірностей потрапляння вибірових значень до  $X_i$ , які відповідно дорівнюють 0,237; 0,500; 0,263.

Обчислимо значення відхилення Пірсона між емпіричним і гіпотетичним розподілами:

$$\begin{aligned} \eta_{\text{зн}} &= \frac{(v_0 - n\bar{p}_0)^2}{n\bar{p}_0} + \frac{(v_1 - n\bar{p}_1)^2}{n\bar{p}_1} + \frac{(v_2 - n\bar{p}_2)^2}{n\bar{p}_2} = \\ &= \frac{(476 - 478,74)^2}{478,74} + \frac{(1017 - 1010)^2}{1010} + \frac{(527 - 531,26)^2}{531,26} = 0,098. \end{aligned}$$

Отже,

$$\eta_{\text{зн}} = 0,098 < 3,84 = t_{0,05; 1} = t_{\alpha; (r-s-1)},$$

тут  $r - 1 - k = 3 - 1 - 1$ ;  $k = 1$  – кількість параметрів, що були оцінені за вибіркою;  $r = 3$  – кількість підмножин, на які поділяється вибіровий простір). Згідно з критерієм  $\chi^2$  гіпотеза про біномний розподіл випадкової величини  $\xi$  не відхиляється. Припущення про те, що кількість хлопчиків у дводітних сім'ях має біномний розподіл, не суперечить даним.

**Приклад 4.1.3.** Спостереження деякої випадкової величини наведені в інтервальному варіаційному ряді

Таблиця 4.2

інтервали	$[-4; 0)$	$[0; 2)$	$[2; 4)$	$[4; 6)$
$n_i$	20	40	30	10

Перевірити гіпотезу про нормальний розподіл генеральної сукупності.

**Розв'язання.** За реалізацією вибірки знайдемо значення найкращих оцінок параметрів гауссівського розподілу: вибіркового середнього  $\bar{x} = 1,4$  та виправленої вибіркової дисперсії  $\tilde{D}_{\text{зн}} = 4,48$ . Висуваємо гіпотезу  $H_0: \xi \in N(1,4; 4,48)$ . Обчислюємо теоретичні ймовірності, як ймовірності попадання гауссівської випадкової величини в у відповідні проміжки. Наприклад,  $p_1 = P(\xi \in [-4; 0)) \approx 0,2494$ . Додаємо проміжки  $(-\infty; -4)$  та  $[6; +\infty)$ . Відповідно обчислюємо  $p_0 = P(\xi \in (-\infty; -4)) \approx 0,005$  та  $p_5 = P(\xi \in [6; +\infty)) \approx 0,07$ . Перевіряємо умову  $\sum_{i=0}^5 p_i = 1$ , а також умову  $np_i \geq 10$ . Остання умова не виконується для проміжків  $(-\infty; -4)$  та  $[6; +\infty)$ , тож об'єднуємо їх із сусідніми проміжками. Відповідно обчислюємо  $\tilde{p}_1 = p_0 + p_1$  та  $\tilde{p}_4 = p_4 + p_5$ . Знаходимо значення статистики критерію  $\eta_{\text{зн}} \approx 1,66$ . За таблицею розподілу Пірсона з'ясовуємо межу критичної області, взявши  $\alpha = 0,05$ ,  $k = 4 - 2 - 1 = 1$ ,  $t = t_{0,05; 1} = t_{\alpha; k} = 3,84$ . Робимо висновок: на рівні значущості 0,05 дослідні данні не суперечать тому, що дана реалізація взята з гауссівського розподілу.

## 4.2. Застосування критерію $\chi^2$ в схемі Бернуллі

Нехай у послідовності  $n$  незалежних випробувань подія  $A$  відбулася  $m$  разів. Потрібно на рівні значущості  $\alpha$  перевірити гіпотезу, що  $P(A) = p$ . Дані можна розглядати як вибірку з  $n$  значень випадкової величини  $\xi$ , що

набуває значення 0 або 1 в залежності від того відбулася чи ні подія  $A$  в конкретному випробуванні (індикатор появи події  $A$ ).

Висуваємо гіпотезу  $H_0$  подія  $A$  відбувається з ймовірністю  $P(A) = p$ , відповідно, не відбувається – з ймовірністю  $q = 1 - p$ .

Згідно критерію  $\chi^2$ , записуємо статистику

$$\eta = \frac{(m-np)^2}{np} + \frac{(n-m-nq)^2}{nq}.$$

За теоремою Пірсона ця статистика прямує у слабкому розумінні до розподілу  $\chi^2$  (Пірсона) з однією ступінню вільності. Подальша перевірка гіпотези проводиться так само як описано в попередньому параграфі. Тобто, якщо  $\eta_{\text{зн}} < t_{\alpha,1}$ , то на рівні значущості  $\alpha$ , дані не суперечать висунутій гіпотезі  $H_0$ ; якщо ж  $\eta_{\text{зн}} > t_{\alpha,1}$ , то на рівні значущості  $\alpha$ , дані суперечать висунутій гіпотезі  $H_0$ .

### *Приклади розв'язання задач*

**Приклад 4.2.1.** При  $n = 4040$  підкиданнях монети Ж. Бюффон отримав  $n = 2048$  випадань «герба».

Чи узгоджується з цими даними гіпотеза: ймовірність випадань «герба» дорівнює 0,5?

**Розв'язання.** Згідно умови задачі підрахуємо значення статистики Пірсона

$$\eta_{\text{зн}} = \frac{\left(\frac{2048 - \frac{4040}{2}}{\frac{4040}{2}}\right)^2}{\frac{4040}{2}} + \frac{\left(\frac{1992 - \frac{4040}{2}}{\frac{4040}{2}}\right)^2}{\frac{4040}{2}} = 0,776.$$

За таблицею розподілу Пірсона з'ясуємо межу критичної області, взявши  $\alpha = 0,05$ ,  $k = 1$ ,  $t = t_{0,05;1} \approx 3,84$ . Робимо висновок: на рівні значущості 0,05 дослідні данні не суперечать тому, що ймовірність випадання герба дорівнює 0,5.

### 4.3. Критерій незалежності $\chi^2$

Критерій  $\chi^2$  дає змогу перевірити гіпотезу про незалежність двох випадкових величин  $\xi$  та  $\eta$ .

Статистикою критерію є випадкова величина

$$\eta = \sum_{i=1}^l \sum_{j=1}^k \frac{(v_{ij} - m_{ij})^2}{m_{ij}},$$

де  $v_{ij}$  - кількість випадків, коли одночасно спостерігалися  $\xi = x_i$  та  $\eta = y_j$ ,  $i = \overline{1, l}$ ,  $j = \overline{1, k}$  (для неперервних випадкових величин  $i$  та  $j$  - номери відповідних інтервалів),

$$m_{ij} = \frac{v_{i \cdot} \cdot v_{\cdot j}}{n}, \quad v_{i \cdot} = \sum_{j=1}^k v_{ij}, \quad v_{\cdot j} = \sum_{i=1}^l v_{ij};$$

де  $l$  і  $k$  - кількість значень, яких набувають випадкові величини  $\xi$  та  $\eta$ ;  $n = \sum_{i=1}^l \sum_{j=1}^k v_{ij}$  - обсяг вибірки.

Вибір табличного значення  $t_{\alpha; (l-1)(k-1)}$  із таблиці Пірсона та прийняття рішення проводиться аналогічно описаній вище процедурі.

### Приклади розв'язання задач

**Приклад 4.3.1.** Проведено 300 спостережень одночасно над випадковими величинами  $\xi$  та  $\eta$ , які набувають значень 1, 2 та 1, 2, 3 відповідно. Кількості пар  $v_{ij}$ , які спостерігалися наведено у таблиці

$\xi$	$\eta$			$v_{i\cdot}$
	1	2	3	
1	32	68	50	150
2	40	70	40	150
$v_{\cdot j}$	72	138	90	300

Перевірити гіпотезу за критерієм  $\chi^2$  про незалежність випадкових величин  $\xi$  та  $\eta$  на рівні значущості 0,01.

**Розв'язання.** Знайдемо величини  $m_{ij} = \frac{v_{i\cdot} \cdot v_{\cdot j}}{n}$ . Матриця, що складена з цих величин буде такою  $\begin{pmatrix} 36 & 69 & 45 \\ 36 & 69 & 45 \end{pmatrix}$ , а матриця, елементами, якої є  $(v_{ij} - m_{ij})^2 - \begin{pmatrix} 16 & 1 & 25 \\ 16 & 1 & 25 \end{pmatrix}$ .

Далі складемо матрицю, елементами якої будуть величини  $\frac{(v_{ij} - m_{ij})^2}{m_{ij}}$ . Вона буде такою  $\begin{pmatrix} 0,44 & 0,014 & 0,55 \\ 0,44 & 0,014 & 0,55 \end{pmatrix}$ . Підсумувавши елементи цієї матриці знаходимо значення статистики критерію  $\eta_{\text{зн}} = 2,008$ . Кількість ступенів вільності – 2. Межу критичної області знаходимо за таблицею Пірсона  $t_{0,01; 2} = 9,2$ . Робимо висновок: на рівні значущості 0,01 дослідні данні не суперечать тому, що випадкові величини  $\xi$  та  $\eta$  є незалежними.

## Задачі для самостійного розв'язання

### 4.1. Дискретний варіаційний ряд деякої генеральної сукупності має

вигляд:

а)

варіанти	0	1	2	3	4
$n_i$	110	65	21	3	1

б)

варіанти	0	1	2	3	4	5	6	7
$n_i$	112	168	130	68	32	5	1	1

в)

варіанти	0	1	2	3	4	5
$n_i$	229	211	93	35	7	1

На рівні значущості 0,05 за критерієм  $\chi^2$  перевірити гіпотезу про пуассонівський розподіл генеральної сукупності.

### 4.2. Інтервальний варіаційний ряд деякої генеральної сукупності має

вигляд:

а)

інтервали	[0; 5)	[5; 10)	[10; 15)	[15; 20)	[20; 25)
$n_i$	15	75	100	50	10

б)

інтервали	$[-3; -1)$	$[-1; 0)$	$[0; 1)$	$[1; 2)$	$[2; 3)$	$[3; 5)$
$n_i$	13	15	24	25	13	10

в)

інтервали	$[-4; 0)$	$[0; 2)$	$[2; 4)$	$[4; 6)$
$n_i$	20	40	30	10

На рівні значущості 0,05 за критерієм  $\chi^2$  перевірити гіпотезу про нормальний розподіл генеральної сукупності.

**4.3.** За часи другої світової війни на Лондон впало 537 снарядів. Уся територія Лондона була поділена на 576 ділянок площею 0,25 км<sup>2</sup>. В таблиці наведено кількість ділянок  $n_k$ , на які впало  $k$  снарядів

$k$	0	1	2	3	4	5 і більше
$n_k$	229	211	93	35	7	1

Чи узгоджуються ці данні з гіпотезою про пуассонівський розподіл кількості снарядів, що впали на кожну ділянку?

**4.4.** Кількість випадінь «герба» при 20 підкиданнях двох однакових монет розподілилася наступним чином

Кількість «гербів»	0	1	2
Кількість підкидань	4	8	8

Чи узгоджується цей результат із припущенням про симетричність монет та незалежності результатів підкидань?

**4.5.** На іспиті студент відповідає тільки на одне питання по одній із трьох частин курсу з «Теорії ймовірностей». Аналіз питань, що були задані 60 студентам, показав, що 23 студента отримали питання з першої частини курсу, 15 – з другої і 22 – з третьої частини.

Чи можна вважати, що студент отримає з рівною ймовірністю питання по будь-якій частині курсу? Гіпотезу перевірити на рівні значущості 0,1.

**4.6.** Нижче наведені результати опитування 100 студентів перших трьох курсів на питання «Чи вважаєте Ви, що паління заважає навчанню?»

відповідь	I курс	II курс	III курс
ні	15	10	-
не знаю	8	5	7
так	-	30	25

Чи підтверджують ці данні припущення про те, що відношення до паління студентів різних курсів є різним? Гіпотезу перевірити на рівні значущості 0,01.

**4.7.** Двовимірний дискретний випадковий вектор може набувати чотири значення:  $(0; 0)$ ,  $(0; 1)$ ,  $(1; 0)$ ,  $(1; 1)$ . 180 незалежних спостережень дали такі результати:

$(0; 0)$  спостерігалось 39 разів;  $(0; 1)$  – 50 разів;  $(1; 0)$  – 53 рази;  $(1; 1)$  – 38 разів. Чи можна вважати координати цього вектору незалежними? Гіпотезу перевірити на рівні значущості 0,1.

**4.8.** Абітурієнти, які вступають до інституту, були навмання розподілені на два потоки по 300 чоловік у кожному. Результати екзамену з однієї і тієї самої дисципліни в кожному потоці виявилися такими:

1-й потік: бали «2», «3», «4» та «5» одержали відповідно 33, 43, 80 та 144 чоловіки;

2-й потік: бали «2», «3», «4» та «5» одержали відповідно 39, 35, 72 та 154 чоловіки.

Чи існує зв'язок між оцінкою на екзамені й потоком, в якому абітурієнт складав екзамен?

**4.9.** Чотири монети були підкинуті 20 160 разів, при цьому комбінації: чотири «герби», три «герби» і «решітка», два «герби» і дві «решітки», один «герб» і три «решітки», чотири «решітки» з'явилися відповідно таку кількість разів: 1181, 4909, 7583, 5085, 1402 (дані експерименту, проведеного В.І. Романовським).

Чи свідчать ці дані про те, що кількість «гербів», яка з'явилася на чотирьох монетах, є біномно розподіленою випадковою величиною?

Сформулювати і розв'язати поставлену задачу як задачу перевірки статистичних гіпотез.

**4.10.** Помилкове з'єднання телефонних номерів.

У таблиці наведено дані про кількість помилкових з'єднань телефонних номерів. Усього спостерігалось 267 номерів.

$k$	$n_k$	$k$	$n_k$	$k$	$n_k$
0	0	6	22	12	18
1	1	7	43	13	12
2	0	8	31	14	7
3	5	9	40	15	6
4	11	10	35	16	2
5	14	11	20	Разом	267

Тут  $k$  – кількість помилкових з'єднань;  $n_k$  – кількість номерів, на яких було зафіксовано точно  $k$  помилкових з'єднань,  $k = 0, 1, \dots, 16$ .

Перевірити гіпотезу про пуассонівський розподіл кількості помилкових з'єднань.

**4.11.** При 24 000 підкиданнях монети К. Пірсон отримав 12 012 випадів «герба».

Чи узгоджується гіпотеза про симетричність монет з цими даними?

**4.12.** В одному з експериментів із гральними кубиками Уелдон підкинув кубики 49 152 рази. При цьому в 25 145 випадках випадали числа 4, 5 або 6.

Чи узгоджується з цими даними гіпотеза про симетричність кубиків?

**4.13.** Нижче в таблиці наведено дані про появу цифр 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 серед перших 800 знаків числа  $\pi$ .

$k$	0	1	2	3	4	5	6	7	8	9	Разом
$n_k$	74	92	83	79	80	73	77	75	76	91	800

Тут  $k$  – цифра в десятковому запису числа  $\pi$ ;  $n_k$  – кількість появ цифри  $k$  серед перших 800 знаків числа  $\pi$ .

Чи узгоджується гіпотеза про рівномірність появ кожної цифри в десятковому запису числа  $\pi$  з наведеними даними?

**4.14.** При 120 киданнях гральної кістки шість очок випало 40 разів. Чи узгоджується з цими даними гіпотеза: ймовірність випадіння «шістки» дорівнює  $\frac{1}{6}$ ?

**4.15.** Проведено 200 спостережень одночасно над випадковими величинами  $\xi$  та  $\eta$ , які набувають значень 1, 2 та 1, 2, 3 відповідно. Кількості пар  $v_{ij}$ , що спостерігалися наведено у таблиці.

Перевірити гіпотезу про незалежність випадкових величин  $\xi$  та  $\eta$  на рівні значущості 0,05.

$\xi$	$\eta$			
	1	2	3	$v_{i\cdot}$
1	25	50	25	100
2	52	41	7	100
$v_{\cdot j}$	77	91	32	200

**4.16.** В таблиці наведено результати 145 спостережень двовимірного дискретного випадкового вектору  $(\xi, \eta)$ . Перевірити гіпотезу про незалежність випадкових величин  $\xi$  та  $\eta$  на рівні значущості 0,05.

$\xi$	$\eta$			$v_{i.}$
	3	4	5	
0	45	25	15	85
1	11	11	13	35
2	9	9	7	25
$v_{.j}$	65	45	35	145

## Глава 5.      Перевірка гіпотез про параметри розподілу генеральної сукупності

### 5.1. Перевірка простих гіпотез про параметри розподілу генеральної сукупності

Методика перевірки простої параметричної гіпотези  $H_0$  проти простої конкуруючої гіпотези  $H_1$  ґрунтується на тому, що критичну область  $G \setminus W$  слід вибрати таким чином, щоб ймовірність попадання в неї статистики критерію була мінімальною і дорівнювала рівню значущості  $\alpha$ , якщо дослідні дані не суперечать нульовій гіпотезі  $H_0$  та максимальною у протилежному випадку. Тобто критична область повинна бути такою, щоб при заданому рівні значущості  $\alpha$  потужність критерію  $1 - \beta$  була максимальною. Назвемо **найкращою критичною областю** таку множину, що забезпечує максимальну потужність критерію. Побудова такої області ґрунтується на **лемі Неймана-Пірсона**.

Серед всіх критеріїв заданого рівня значущості  $\alpha$ , які перевіряють просту параметричну гіпотезу  $H_0: \theta = \theta_0$  проти простої конкуруючої гіпотези  $H_1: \theta = \theta_1$  критерій відношення правдоподібності є найбільш потужним, тобто найкраща критична область складається з таких точок вибіркового простору, які попадають в множину

$$W_{C_\alpha} = \left\{ \vec{x} \in \mathbb{R}^n : \frac{L_{H_1}(\vec{x})}{L_{H_0}(\vec{x})} \geq C_\alpha \right\},$$

де у чисельнику стоїть функція правдоподібності, в припущенні, що справджується гіпотеза  $H_1$ , у знаменнику – відповідно гіпотеза  $H_0$ . Стала  $C_\alpha$  вибирається з умови

$$P \left\{ \frac{L_{H_1}(\vec{\xi})}{L_{H_0}(\vec{\xi})} \geq C_\alpha \right\} = P \left\{ \ln \frac{L_{H_1}(\vec{\xi})}{L_{H_0}(\vec{\xi})} \geq \ln C_\alpha \right\} = \alpha.$$

Знайшовши найкращу критичну область, можна обчислити помилку другого роду  $\beta$  з умови  $P \left\{ \ln \frac{L_{H_0}(\vec{\xi})}{L_{H_1}(\vec{\xi})} \leq \ln C_\alpha \right\} = \beta$ , а також визначити мінімальний обсяг вибірки, при якому досягаються наперед задані помилки першого та другого роду, тобто  $\alpha$  та  $\beta$ . Для цього необхідно розв'язати систему

$$\begin{cases} P \left\{ \ln \frac{L_{H_1}(\vec{\xi})}{L_{H_0}(\vec{\xi})} \geq \ln C_\alpha \right\} = \alpha; \\ P \left\{ \ln \frac{L_{H_0}(\vec{\xi})}{L_{H_1}(\vec{\xi})} \leq \ln C_\alpha \right\} = \beta. \end{cases}$$

### **Приклади розв'язання задач**

**Приклад 5.1.1.** Побудувати найкращу критичну область за лемою Неймана-Пірсона для параметра  $\mu$  нормального закону розподілу з відомою дисперсією  $\sigma^2$ . Введемо дві прості гіпотези (основну та альтернативну):  $H_0: \mu = \mu_0$ ,  $H_1: \mu = \mu_1$ , де  $\mu_0$  і  $\mu_1$  – деякі задані значення, пов'язані, наприклад, нерівністю  $\mu_0 < \mu_1$ .

У розглянутому випадку функція правдоподібності має вигляд:

$$L(\xi_1, \dots, \xi_n; \mu) = \left( \frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^n (\xi_i - \mu)^2 \right),$$

а відношення правдоподібності:

$$\begin{aligned}\varphi(\vec{\xi}) &= \frac{L(\xi_1, \dots, \xi_n; \mu_1)}{L(\xi_1, \dots, \xi_n; \mu_0)} = \\ &= \exp\left(\frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1}^n \xi_i\right) \exp\left(-\frac{n(\mu_1 - \mu_0)^2}{2\sigma^2}\right).\end{aligned}$$

У даному випадку нерівність

$$\varphi(\vec{\xi}) = \exp\left(\frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1}^n \xi_i\right) \exp\left(-\frac{n(\mu_1 - \mu_0)^2}{2\sigma^2}\right) \geq C_\varphi$$

рівносильна нерівності

$$\sum_{i=1}^n \xi_i \geq C, \quad (5.1)$$

де константу  $C$  обирають з умови забезпечення заданого рівня значущості  $\alpha$ :

$$P(\sum_{i=1}^n \xi_i \geq C) = \alpha. \quad (5.2)$$

Дійсно,

$$\begin{aligned}\ln\left(\exp\left(\frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1}^n \xi_i\right) \exp\left(-\frac{n(\mu_1 - \mu_0)^2}{2\sigma^2}\right)\right) &= \\ &= \frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=1}^n x_i - \frac{n(\mu_1 - \mu_0)^2}{2\sigma^2} \geq \ln C_\varphi,\end{aligned}$$

звідки випливає, що

$$\sum_{i=1}^n x_i \geq \frac{\sigma^2}{\mu_1 - \mu_0} \left(\ln C_\varphi - \frac{n(\mu_1 - \mu_0)^2}{2\sigma^2}\right) = C.$$

Випадкова величина  $\sum_{i=1}^n \xi_i$  має нормальний розподіл з математичним сподіванням  $n\mu$  і дисперсією  $n\sigma^2$ . Тому умову (5.2) можна записати у вигляді

$$0,5 - \Phi\left(\frac{C - n\mu_0}{\sigma\sqrt{n}}\right) = \alpha, \quad (5.3)$$

Використовуючи таблицю значень функції Лапласа, знаходимо

$$\frac{C - n\mu_0}{\sigma\sqrt{n}} = t.$$

Таким чином, константа  $C$ , що задає критичну область в (5.1), визначається рівністю

$$C = n\mu_0 + t\sigma\sqrt{n}. \quad (5.4)$$

При цьому ймовірність зробити помилку другого роду

$$\beta = P\{\sum_{i=1}^n \xi_i \leq C \mid \mu = \mu_1\} = 0,5 + \Phi\left(\frac{C - n\mu_1}{\sigma\sqrt{n}}\right) \quad (5.5)$$

є мінімально можливою при даному значенні  $\alpha$ .

**Приклад 5.1.2.** Якщо за умовою прикладу 5.1.1 нерівність  $\mu_0 < \mu_1$  замінити нерівністю  $\mu_1 < \mu_0$ , то в цьому випадку критична множина задається нерівністю

$$\sum_{i=1}^n x_i \leq C,$$

де константу  $C$  вибирають з умови

$$P\{\sum_{i=1}^n \xi_i \leq C \mid \mu = \mu_0\} = \alpha.$$

Таким чином,

$$0,5 + \Phi\left(\frac{C - n\mu_0}{\sigma\sqrt{n}}\right) = \alpha.$$

З останньої рівності знаходимо  $C$ .

**Приклад 5.1.3.** Побудувати найкращу критичну область за лемою Неймана-Пірсона у випадку експоненційного розподілу з параметром  $\lambda$ . Введемо дві прості гіпотези:

$$H_0: \lambda = \lambda_0, H_1: \lambda = \lambda_1,$$

де  $\lambda_0 < \lambda_1$ . У цьому випадку функція правдоподібності

$$L(\xi_1, \dots, \xi_n; \lambda) = \lambda^n \exp(-\lambda \sum_{i=1}^n \xi_i).$$

Таким чином,

$$\varphi(\vec{\xi}) = \left(\frac{\lambda_1}{\lambda_0}\right)^n \exp(-(\lambda_1 - \lambda_0) \sum_{i=1}^n \xi_i).$$

Звідси видно, що критичну множину можна задати нерівністю

$$\sum_{i=1}^n x_i \leq C,$$

де константу  $C$  вибирають з умови

$$P\{\sum_{i=1}^n (\xi_i \leq C \mid \lambda = \lambda_0)\} = \alpha.$$

Випадкова величина  $2\lambda(\xi_1 + \dots + \xi_n)$  при  $\lambda = \lambda_0$  у слабкому сенсі прямує до випадкової величини, яка має  $\chi^2$  розподіл з  $2n$  ступенями вільності. Виходячи з цього, отримуємо вираз для константи  $C$ :

$$C = \frac{\chi_{\alpha, 2n}^2}{2} \lambda_0,$$

де  $\chi_{\alpha, 2n}^2$  квантиль рівня  $\alpha$  для  $\chi^2$  розподілу з  $2n$  ступенями вільності. При цьому ймовірність зробити помилку другого роду дорівнює

$$\beta = P\{\sum_{i=1}^n \xi_i > C \mid \lambda = \lambda_1\} = 1 - H_{2n}(2\lambda_1 C) = 1 - H_{2n}\left(\chi_{\alpha, 2n}^2 \frac{\lambda_1}{\lambda_0}\right),$$

де  $H_{2n}(t)$  – функція розподілу випадкової величини, що має  $\chi^2$  розподіл з  $2n$  ступенями вільності.

**Приклад 5.1.4.** Побудувати найкращу критичну область за лемою Неймана-Пірсона для параметра біномного розподілу проведемо для випадку двох простих гіпотез:

$$H_0: p = p_0, H_1: p = p_1,$$

де  $p$  – ймовірність «успіху» в одному випробуванні при реалізації схеми незалежних випробувань Бернуллі, а  $p_0$  і  $p_1$  – задані значення параметра, що задовольняють нерівності  $p_0 < p_1$ .

Нехай кількість випробувань достатньо велика і  $\xi_j$  – результат  $j$ -го випробування. Випадкова величина  $\xi_j$  приймає значення 0 і 1 з ймовірностями  $1 - p$  і  $p$  відповідно. Функція правдоподібності у цьому випадку має вигляд

$$L(\xi_1, \dots, \xi_n; p) = C_n^{k(\vec{\xi})} p^{k(\vec{\xi})} (1 - p)^{n - k(\vec{\xi})},$$

де  $k(\vec{\xi}) = \sum_{i=1}^n \xi_i$  – загальна кількість «успіхів» у серії з  $n$  випробувань.

Відношення правдоподібності визначається рівністю

$$\varphi(\vec{\xi}) = \frac{L(\xi_1, \dots, \xi_n; \mu_1)}{L(\xi_1, \dots, \xi_n; \mu_0)} = \left(\frac{p_1}{p_0}\right)^{k(\vec{\xi})} \left(\frac{1-p_0}{1-p_1}\right)^{n-k(\vec{\xi})}.$$

Отже, критична множина для оптимального критерію Неймана-Пірсона у даному випадку має вигляд

$$k(\vec{x}) = \sum_{i=1}^n x_i \geq C. \quad (5.6)$$

Константу  $C$  вибирають виходячи з умови

$$P\{\sum_{i=1}^n (\xi_i \geq C \mid p = p_0)\} = \alpha.$$

Розподіл випадкової величини  $k(\vec{\xi})$  при достатньо великих  $n$  у відповідності з відомою інтегральною теоремою Муавра-Лапласа має асимптотично нормальний розподіл з математичним сподівання  $\mu = np$  і дисперсією  $\sigma^2 = np(1 - p)$ . Використовуючи вказану умову, виберемо константу  $C$  в (5.6) з умови забезпечення заданої умови рівня значущості  $\alpha$ , тобто з умови

$$P\{(k(\vec{\xi}) \geq C \mid p = p_0)\} \approx 0,5 - \Phi\left(\frac{C - np_0}{\sqrt{np_0(1 - np_0)}}\right) = \alpha, \quad (5.7)$$

## 5.2. Перевірка складних гіпотез про параметри розподілу генеральної сукупності

Розглядаємо задачу перевірки простої основної гіпотези  $H_0: \theta = \theta_0$  проти складної конкуруючої гіпотези  $H_1: \theta = \theta_1 > \theta_0$  або  $H_2: \theta = \theta_1 < \theta_0$ . Побудований на основі відношення правдоподібності критерій перевірки простої гіпотези буде оптимальним і для перевірки простих основних гіпотез проти наведених складних конкуруючих. Побудована критична область у випадку конкуруючої гіпотези  $H_1$  буде правостороння, відповідно, у випадку конкуруючої гіпотези  $H_2$  – лівостороння.

Наведемо приклади перевірки гіпотез про параметри *гауссівської генеральної сукупності*.

*Перевірка гіпотези про числове значення математичного сподівання гауссівської генеральної сукупності при відомій дисперсії.*

Висунемо основну гіпотезу  $H_0: E\xi = a_0$ , конкуруюча гіпотеза  $H_1: E\xi = a_1 > a_0$ .

Найкращою точковою оцінкою математичного сподівання є вибіркове середнє  $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$ , що має гауссівський розподіл з математичним сподіванням  $a_0$  та дисперсією  $\frac{\sigma^2}{n}$ .

Критична область буде знаходитись з умови  $P\left(\frac{1}{n} \sum_{i=1}^n \xi_i \geq C\right) = \alpha$ , вона є відповідно правостороння. При знаходженні  $C$  користуємось таблицею значень функції Лапласа.

***Перевірка гіпотези про числове значення математичного сподівання гауссівської генеральної сукупності при невідомій дисперсії.***

Висунемо основну гіпотезу  $H_0: E\xi = a_0$ , конкуруюча гіпотеза  $H_1: E\xi = a_1 < a_0$ .

Найкращою точковою оцінкою математичного сподівання є вибіркове середнє  $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$ , що має гауссівський розподіл з математичним сподіванням  $a_0$  та дисперсією  $\frac{\sigma^2}{n}$ , але  $\sigma^2$  невідоме, тому статистикою критерію, що перевіряє гіпотезу  $H_0$  буде  $\frac{(\bar{\xi} - a_0)\sqrt{n}}{\sqrt{D\bar{\xi}}}$ , що має розподіл Стюдента з  $n - 1$  ступенями вільності.

Критична область буде знаходитись з умови  $P\left(\frac{(\bar{\xi} - a_0)\sqrt{n}}{\sqrt{D\bar{\xi}}} < C\right) = \alpha$ , вона є відповідно лівостороння. При знаходженні  $C$  користуємось таблицею значень розподілу Стюдента.

***Перевірка гіпотези про числове значення дисперсії гауссівської генеральної сукупності при відомому та невідомому математичному сподіванні.***

Припустимо, що математичне сподівання гауссівської генеральної сукупності відоме. Висунемо основну гіпотезу  $H_0: D\xi = \sigma_0^2$ , конкуруюча гіпотеза  $H_1: D\xi = \sigma_1^2 > \sigma_0^2$ . Найкращою точковою оцінкою дисперсії при

відомому математичному сподіванні є вибіркова дисперсія  $D^*\xi = \frac{1}{n} \sum_{i=1}^n (\xi_i - E\xi)^2$ . Статистика  $\frac{nD^*\xi}{\sigma_0^2}$  має розподіл  $\chi_n^2$ . Саме для побудови критичної області й використовується таблиця  $\chi^2$  розподілу, що наведена у додатку. У випадку, коли математичне сподівання невідоме за оцінку дисперсії беремо виправлену вибірку дисперсію  $\tilde{D}\xi = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2$ . Зазначимо, що статистика  $\frac{(n-1)\tilde{D}\xi}{\sigma_0^2}$  має розподіл  $\chi_{n-1}^2$ .

Якщо точкова оцінка параметра  $\theta$  будь-якої генеральної сукупності є асимптотично нормальною, то для перевірки гіпотези про значення цієї оцінки застосовується гауссівський розподіл.

На практиці часто виникає необхідність порівняти середній результат однієї серії експериментів із середнім результатом іншої серії. Виникає питання: чи можна розбіжність, що виникла пояснити випадковою помилкою експерименту, чи ця розбіжність не випадкова. Так ми приходимо до необхідності перевірити наступні гіпотези.

***Перевірка гіпотези про рівність математичних сподівань двох гауссівських генеральних сукупностей при відомих дисперсіях.***

Маємо дві реалізації обсягами  $n_1$  та  $n_2$  відповідно двох гауссівських генеральних сукупностей  $\xi \sim N\left(E\xi, \frac{\sigma_1^2}{n_1}\right)$  та  $\eta \sim N\left(E\eta, \frac{\sigma_2^2}{n_2}\right)$ . Дисперсії цих генеральних сукупностей відомі. На рівні значущості  $\alpha$  потрібно перевірити просту основну гіпотезу  $H_0: E\xi = E\eta$  проти однієї з конкуруючих гіпотез  $H_1: E\xi > E\eta$  або  $H_2: E\xi < E\eta$ .

Статистикою критерія, що перевіряє гіпотезу  $H_0$  є випадкова величина  $\frac{\bar{\xi} - \bar{\eta}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ , яка, в припущенні що основа гіпотеза вірна, має стандартній гауссівський розподіл  $N(0, 1)$ .

Критична область береться правостороння чи лівостороння, в залежності від конкуруючої гіпотези. Межа критичної області береться з таблиці значень функції Лапласа.

***Перевірка гіпотези про рівність дисперсій двох гауссівських генеральних сукупностей при відомих або невідомих математичних сподіваннях.***

Маємо дві реалізації обсягами  $n_1$  та  $n_2$  відповідно двох гауссівських генеральних сукупностей  $\xi \sim N\left(E\xi, \frac{\sigma_1^2}{n_1}\right)$  та  $\eta \sim N\left(E\eta, \frac{\sigma_2^2}{n_2}\right)$ . Дисперсії цих генеральних сукупностей невідомі. На рівні значущості  $\alpha$  потрібно перевірити просту основну гіпотезу  $H_0: D\xi = D\eta$  проти однієї з конкуруючих гіпотез  $H_1: D\xi > D\eta$  або  $H_2: D\xi < D\eta$ .

Припустимо, що відомі математичні сподівання  $E\xi$  та  $E\eta$ . Розглянемо дві статистики

$$h_1 = \frac{n_1 \cdot D^* \xi}{\sigma_1^2} \sim \chi_{n_1}^2, \quad h_2 = \frac{n_2 \cdot D^* \eta}{\sigma_2^2} = \chi_{n_2}^2.$$

Як відомо, статистика  $\gamma = \frac{h_1/n_1}{h_2/n_2} = \frac{D^* \xi}{D^* \eta} \sim F(n_1, n_2)$ , тобто має розподіл Фішера з  $n_1$  та  $n_2$  ступенями вільності. Критична область береться правостороння чи лівостороння, в залежності від конкуруючої гіпотези. Межа критичної області береться з таблиці значень розподілу Фішера.

Якщо ж математичні сподівання  $E\xi$  та  $E\eta$  невідомі, то розглядаємо статистику

$$h_1 = \frac{(n_1-1) \cdot \bar{D}\xi}{\sigma_1^2} \sim \chi_{n_1-1}^2 \text{ та } h_2 = \frac{(n_2-1) \cdot \bar{D}\eta}{\sigma_2^2} \sim \chi_{n_2-1}^2.$$

Тоді статистика

$$\gamma = \frac{h_1/(n_1-1)}{h_2/(n_2-1)} = \frac{\bar{D}\xi}{\bar{D}\eta} \sim F(n_1 - 1, n_2 - 1),$$

тобто має розподіл Фішера з  $n_1 - 1$  та  $n_2 - 1$  ступенями вільності. Критична область береться правостороння чи лівостороння, в залежності від конкуруючої гіпотези. Межа критичної області береться з таблиці значень розподілу Фішера.

***Перевірка гіпотези про рівність математичних сподівань двох гауссівських генеральних сукупностей при невідомих, але рівних дисперсіях.***

Маємо дві реалізації обсягами  $n_1$  та  $n_2$  відповідно двох гауссівських генеральних сукупностей  $\xi \sim N\left(E\xi, \frac{\sigma^2}{n_1}\right)$  та  $\eta \sim N\left(E\eta, \frac{\sigma^2}{n_2}\right)$ . Дисперсії цих генеральних сукупностей невідомі, але рівні. На рівні значущості  $\alpha$  потрібно перевірити просту основну гіпотезу  $H_0: E\xi = E\eta$  проти однієї з конкуруючих гіпотез  $H_1: E\xi > E\eta$  або  $H_2: E\xi < E\eta$ .

Оскільки дисперсії невідомі, то ми можемо побудувати тільки їх точкові оцінки

$$h_1 = \frac{(n_1-1) \cdot \bar{D}\xi}{\sigma^2} \sim \chi_{n_1-1}^2 \text{ та } h_2 = \frac{(n_2-1) \cdot \bar{D}\eta}{\sigma^2} \sim \chi_{n_2-1}^2.$$

Тоді  $h = h_1 + h_2 \sim \chi_{n_1+n_2-2}^2$ .

Статистика критерію в цьому випадку буде такою

$$\gamma = \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\frac{n_1+n_2}{n_1 n_2} \cdot \frac{(n_1-1)\bar{D}\xi + (n_2-1)\bar{D}\eta}{n_1+n_2-2}}} \sim St_{n_1+n_2-2}.$$

Тобто має розподіл Стюдента з  $n_1 + n_2 - 2$  ступенями вільності. Критична область береться правостороння чи лівостороння, в залежності від конкуруючої гіпотези. Межа критичної області береться з таблиці значень розподілу Стюдента. Гіпотезу про рівність дисперсій слід перевіряти.

***Перевірка гіпотези про рівність ймовірностей біномних генеральних сукупностей при великому обсязі вибірки.***

Нехай є реалізації двох біномних генеральних сукупностей  $\xi$  та  $\eta$  обсягами  $n_1$  та  $n_2$  відповідно, причому обсяги вибірок достатньо великі так, що можна застосувати асимптотичну нормальність оцінки ймовірності «успіху» в одному випробуванні.

Потрібно перевірити гіпотезу  $H_0: p_1 = p_2$ , проти конкуруючої гіпотези  $H_1: p_1 > p_2$  або  $H_2: p_1 < p_2$ . Оскільки обсяги вибірок великі, то за центральною граничною теоремою, вважаємо, що

$$\frac{m_1}{n_1} = p_1^* \approx N\left(p_1, \frac{p_1 \cdot q_1}{n_1}\right), \frac{m_2}{n_2} = p_2^* \approx N\left(p_2, \frac{p_2 \cdot q_2}{n_2}\right).$$

Тоді статистика  $p_1^* - p_2^* \approx N\left(0, \frac{p_1 \cdot q_1}{n_1} + \frac{p_2 \cdot q_2}{n_2}\right)$

Статистика критерію, що перевіряє гіпотезу  $H_0: p_1 = p_2 = p \in$

$$Y = \frac{p_1^* - p_2^*}{\sqrt{p^* q^* \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \approx N(0, 1),$$

де в якості  $p^*$  беремо  $\frac{m_1 + m_2}{n_1 + n_2}$ .

Критична область береться правостороння чи лівостороння, в залежності від конкуруючої гіпотези. Межа критичної області береться з таблиці значень функції Лапласа.

### **Приклади розв'язання задач**

**Приклад 5.2.1.** За результатами 9 замірів встановлено, що середній час виготовлення деталі дорівнює 48 с. Припускаючи, що час виготовлення деталі є гауссівська випадкова величина з дисперсією  $9 \text{ с}^2$ , на рівні значущості  $\alpha = 0,05$ , вирішити, чи можна прийняти 50 с як нормативний час виготовлення деталі?

**Розв'язання.** Нехай випадкова величина  $\xi$  задає час виготовлення деталі. За умовою  $\xi$  має гауссівський розподіл. Потрібно перевірити основну гіпотезу  $H_0: E\xi = 50$ , проти конкуруючої гіпотези  $H_1: E\xi < 50$ , значення вибіркового середнього за умовою дорівнює  $\bar{x} = 48$ . Критична область лівостороння. Статистика, за якою будемо перевіряти гіпотезу має вигляд  $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i \sim N(50, 1)$ . Отже,  $P(\bar{\xi} < t) = \alpha$  або  $0,5 + \Phi\left(\frac{t-50}{1}\right) = 0,05$ .

За таблицею значень функції Лапласа знаходимо  $t - 50 = -1,64$ , отже  $t_{\text{кр}} = 48,36$ . Критична область  $(-\infty; 48,36)$ . Число 48 попадає у

критичну область, отже, на рівні значущості  $\alpha = 0,05$ , не можна прийняти 50 с як нормативний час виготовлення деталі.

**Приклад 5.2.2.** В умовах попередньої задачі, вважати, що дисперсія невідома, але відоме значення виправленої вибіркової дисперсії по результатах вимірів  $(\tilde{D}\xi)_{\text{зн}} = 9$ . Чи можна тепер на рівні значущості  $\alpha = 0,05$  прийняти 50 с як нормативний час виготовлення деталі?

**Розв'язання.** Основна гіпотеза і конкуруюча залишились такими ж. Критична область лівостороння. Але статистикою критерію буде 
$$h = \frac{(\bar{\xi} - a_0)\sqrt{n}}{\sqrt{\tilde{D}\xi}} \sim St_{n-1}.$$

За таблицею значень розподілу Стюдента знаходимо  $t$ , що відповідає  $\alpha = 0,05$  та  $n - 1 = 8$ . Таким числом буде  $t = -1,86$ , бо критична область лівостороння. Обчислимо значення статистики критерію за умовою задачі  $h_{\text{зн}} = \frac{(48-50)\sqrt{9}}{\sqrt{9}} = -2$ . Оскільки число  $-2$  попадає в критичну область  $(-\infty; -1,86)$ , то на рівні значущості  $\alpha = 0,05$ , не можна прийняти 50 с як нормативний час виготовлення деталі.

**Приклад 5.2.3.** Точність роботи верстата перевіряється по дисперсії контролюючого розміру виробу, який не повинен перевищувати 0,15. За пробою з 25 випадково відібраних деталей обчислена оцінка дисперсії  $(\tilde{D}\xi)_{\text{зн}} = 0,25$ . На рівні значущості  $\alpha = 0,05$  дати відповідь, чи забезпечує верстат потрібну точність. Вважати, що контролюючий розмір виробу є гауссівська випадкова величина.

**Розв'язання.** Верстат буде забезпечувати потрібну точність, якщо дисперсія контролюючого розміру виробу  $\sigma^2 \leq 0,15$ . Оскільки значення виправленої вибіркової дисперсії 0,25, то формулюємо основну гіпотезу  $H_0: D\xi = 0,15$ , а конкуруюча гіпотеза  $H_1: D\xi > 0,15$ . Отже маємо правосторонню критичну область. Статистикою критерію є випадкова величина  $h = \frac{(n-1)\bar{D}\xi}{\sigma_0^2} \sim \chi_{n-1}^2$ . Отже, межу критичної області знаходимо по таблиці  $\chi^2$ , взявши кількість ступенів вільності 24 та  $\alpha = 0,05$ . Отримали  $t_{кр} = 36,42$ , таким чином, критична область  $(36,42; +\infty)$ . Обчислимо значення статистики критерію  $h_{зн} = \frac{24 \cdot 0,25}{0,15} = 40$ . Оскільки значення статистики критерію попало у критичну область, то дослідні дані на рівні значущості  $\alpha = 0,05$  суперечать гіпотезі  $H_0: D\xi = 0,15$ , тобто верстат не забезпечує потрібну точність.

**Приклад 5.2.4.** Витрати сировини на одиницю продукції за старою технологією склали

Витрати сировини	304	307	308
Кількість виробів	1	4	4

За новою технологією

Витрати сировини	303	304	306	308
Кількість виробів	2	6	4	1

Припустивши, що відповідні генеральні сукупності мають гауссівський розподіл, на рівні значущості 0,05 відповісти на питання: чи дає нова технологія економію середньої витрати сировини на один виріб?

**Розв'язання.** Знайдемо за дослідними даними значення незміщених оцінок математичного сподівання та дисперсії обох генеральних сукупностей

$$\bar{x} = \frac{1}{n} \sum_{i=1}^3 x_i n_i = 307,11, (\tilde{D}\xi)_{\text{зн}} = \frac{1}{n-1} \sum_{i=1}^3 (x_i - \bar{x})^2 n_i = 2,378.$$

$$\text{Аналогічно } \bar{y} = 304,77, (\tilde{D}\eta)_{\text{зн}} = 1,685.$$

Перш за все на рівні значущості  $\alpha = 0,05$  перевіримо гіпотезу про рівність дисперсій двох генеральних сукупностей:  $H_0: D\xi = D\eta$  проти конкуруючої гіпотези  $H_1: D\xi > D\eta$ , бо  $(\tilde{D}\xi)_{\text{зн}} > (\tilde{D}\eta)_{\text{зн}}$ . Критична область правостороння, межу критичної області знаходимо за розподілом Фішера, взявши кількість ступенів вільності 8 та 12 відповідно. За таблицею розподілу Фішера  $t_{\text{кр}} = 2,85$ , отже, критичною областю буде інтервал  $(2,85; +\infty)$ . Обчислимо значення статистики критерію  $\gamma = \frac{\tilde{D}\xi}{\tilde{D}\eta}$ .

$$\gamma_{\text{зн}} = \frac{2,378}{1,685} \approx 1,412.$$

Значення статистики критерію потрапило в область прийняття гіпотези, отже, на рівні значущості  $\alpha = 0,05$  дослідні дані не суперечать тому, що генеральні сукупності мають однакові дисперсії.

Тепер висуваємо основну гіпотезу  $H_0: E\xi = E\eta$  проти конкуруючої гіпотези  $H_1: E\xi > E\eta$ , бо  $\bar{x} > \bar{y}$ . Статистикою, що перевіряє цю гіпотезою буде

$$\gamma = \frac{\bar{\xi} - \bar{\eta}}{\sqrt{\frac{n_1+n_2}{n_1 n_2} \cdot \frac{(n_1-1)\bar{D}\xi + (n_2-1)\bar{D}\eta}{n_1+n_2-2}}} \sim St_{n_1+n_2-2}.$$

Обчислимо її значення, підставляючи дослідні дані  $\gamma_{\text{зн}} \approx 3,696$ .

Знайдемо критичну область по таблиці Стьюдента, взявши  $\alpha = 0,05$  і кількість ступенів вільності 20. Маємо  $t_{\text{кр}} = 1,725$ . Отже, критичною областю буде інтервал  $(1,725; +\infty)$ . Значення статистики критерію потрапило у критичну область, отже, на рівні значущості  $\alpha = 0,05$  можна вважати, що нова технологія дає економію середньої витрати сировини на один виріб.

**Приклад 5.2.5.** Порівняти долю браку на рівні значущості  $\alpha = 0,1$  двох партій виробів, маючи наступні дані

Номер партії	Обсяг вибірки	Кількість бракованих виробів у вибірці	Частість появи бракованого виробу
1	$n_1 = 200$	$m_1 = 5$	$(p_1^*)_{\text{зн}} = \frac{m_1}{n_1} = \frac{1}{40}$
2	$n_2 = 300$	$m_2 = 10$	$(p_2^*)_{\text{зн}} = \frac{m_2}{n_2} = \frac{1}{30}$

**Розв'язання.** Потрібно перевірити гіпотезу  $H_0: p_1 = p_2$ , проти конкуруючої гіпотези  $H_2: p_1 < p_2$ , де  $p_1$  та  $p_2$  – ймовірності виготовлення

бракованого виробу відповідно у першій та у другій партіях. Оскільки обсяги обох вибірок великі, то статистикою критерія буде

$$\gamma = \frac{p_1^* - p_2^*}{\sqrt{p^* q^* \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \approx N(0, 1),$$

де в якості  $p^*$  беремо  $\frac{m_1 + m_2}{n_1 + n_2}$ .

Числове значення статистики критерію відповідно дорівнює  $\gamma_{\text{зн}} = -0,54$ . Знайдемо критичну область, для цього скористаємося таблицею функції Лапласа:  $t_{\text{кр}} = -1,28$ . Тобто критичною областю буде інтервал  $(-\infty; -1,28)$ . Значення статистики критерію не потрапляє в критичну область, тому на рівні значущості  $\alpha = 0,1$ , гіпотеза про однакову долю браку в обох партіях не відхиляється.

## Задачі для самостійного розв'язання

**5.1.** Велика партія виробів може мати деякі браковані вироби. Постачальник стверджує, що бракованих виробів 5%, а покупець припускає, що бракованих виробів 10%. Умови поставки наступні: із партії навмання береться та перевіряється 10 виробів. Партія приймається на умовах постачальника, якщо при перевірці виявлено не більше одного бракованого виробу, в протилежному випадку партія приймається на умовах покупця. Сформулювати задачу в термінах перевірки статистичних гіпотез і відповісти на питання:

а) якою є статистика критерію, область її значення, критична область?

б) який розподіл має статистика критерію?

- в) у чому полягає основна та конкуруюча гіпотези?
- г) які ймовірності помилок першого та другого роду?

**5.2.** Перевірка роботи деякого пристрою здійснюється спеціальним тестом. Якщо пристрій працює правильно, то ймовірність проходження тесту дорівнює 0,99, якщо ні, то ймовірність проходження тесту дорівнює 0,4. Пристрій запускається в роботу, якщо тест пройдено 5 разів підряд. Припускаючи, що кількість проходжень тесту пристроєм має біноміальний розподіл, відповісти на питання:

- а) яка критична область статистики критерію та який розподіл вона має?
- б) сформулювати основну гіпотезу, якщо помилка першого роду складається з відхилення правильно працюючого пристрою;
- в) якою є конкуруюча гіпотеза та помилка другого роду?
- г) які ймовірності помилок першого та другого роду?

**5.3.** Із продукції автомата, який виробляє деякі деталі з номінальним значенням розміру, що контролюється  $\mu_0 = 40$  мм, взято вибірку обсягу  $n = 36$ . Значення вибіркового середнього розміру, що контролюється  $\bar{x} = 40,2$  мм. Можна припускати, що фактичні розміри утворюють нормальну генеральну сукупність з дисперсією  $\sigma^2 = 1$  мм<sup>2</sup>. З'ясуйте:

- а) чи можна за результатами проведеного вибіркового дослідження стверджувати, що розмір, який контролюється не більший номінального (прийняти  $\alpha = 0,01$ )?
- б) яка критична множина у цьому випадку?

**5.4.** У відповідності з технічними умовами середній час безвідмовної роботи приладів з великої партії повинен складати не менше 1000 годин з

середнім квадратичним відхиленням 100 годин. Значення вибіркового середнього часу безвідмовної роботи для випадково відібраних 25 приладів виявилось рівним 970 годин. Припустимо, що середнє квадратичне часу безвідмовної роботи для приладів у вибірці співпадає з середнім квадратичним у всій партії, а характеристика що контролюється має нормальний розподіл. З'ясуйте, чи можна вважати, що вся партія приладів не задовольняє технічним умовам, якщо: а)  $\alpha = 0,1$ ; б)  $\alpha = 0,01$ ?

**5.5.** Розв'яжіть попередню задачу за умови, що найкраща оцінка середнього квадратичного відхилення часу безвідмовної роботи, обчислене за вибіркою, дорівнює 115 годин та при: а)  $\alpha = 0,05$ ; б)  $\alpha = 0,01$ .

**5.6.** Стверджується, що кульки, що вироблені верстатом-автоматом, мають середній діаметр  $d_0 = 10$  мм. На рівні значущості  $\alpha = 0,05$ , перевірте цю гіпотезу, якщо у вибірці з  $n = 16$  кульок середній діаметр виявився рівним 10,3 мм, вважаючи, що: а) дисперсія  $\sigma^2$  відома і дорівнює  $\sigma^2 = 1$  мм<sup>2</sup>; б) значення оцінки дисперсії, яке визначено за вибіркою складає  $(\tilde{D}\xi)_{\text{зн}} = 1,21$  мм<sup>2</sup>. Розмір що контролюється має нормальний розподіл.

**5.7.** Знайти потужність критерію в задачі 5.6, вважаючи, що дисперсія  $\sigma^2$  відома і дорівнює  $\sigma^2 = 1$  мм<sup>2</sup> для конкуруючих гіпотез  $H_1: E\xi = 10,1$ ;  $H_2: E\xi = 10,2$ ;  $H_3: E\xi = 10,3$ , якщо а) обсяг вибірки 100; б) обсяг вибірки 16.

**5.8.** Хронометраж затрат часу на зборку вузла машини 20 слюсарів показав, що середній час зборки  $\bar{x} = 77$  хв., а значення виправленої

вибіркової дисперсії дорівнює  $(\tilde{D}\xi)_{\text{зн}} = 4 \text{ хв}^2$ . В припущенні того, що генеральна сукупність має гауссівський розподіл вирішити питання, чи можна на рівні значущості  $\alpha = 0,01$  вважати 80 хв. нормативом трудомісткості?

**5.9.** Із великої партії резисторів одного типу і номіналу випадковим чином відібрано 36 штук. Значення вибіркового середнього величини опору при цьому виявилось рівним 9,3 кОм. На рівні значущості  $\alpha = 0,05$ , перевірити гіпотезу про те, що вибірка була взята із партії з номінальним значенням 10 кОм, якщо: а) дисперсія випадкової величини, що розглядається відома і рівна 4 кОм<sup>2</sup>; б) дисперсія значення опору невідома, а значення виправленої вибіркової дисперсії рівне 6,25 кОм<sup>2</sup>. Розподіл ознаки, що контролюється - гауссівський.

**5.10.** Знайти потужність критерію в задачі 5.9, якщо конкуруюча гіпотеза передбачає номінал партії 9,5 кОм. Який обсяг вибірки необхідно взяти, щоб помилка другого роду дорівнювала 0,01?

**5.11.** Установа має середню продуктивність 1000 кг речовини за добу з середнім квадратичним відхиленням, яке дорівнює 80 кг<sup>2</sup>. При зміні технології продуктивність зростає до 1100 кг речовини за добу з тим самим середнім квадратичним відхиленням. Чи можна вважати, що нова технологія забезпечить підвищення продуктивності, якщо: а)  $\alpha = 0,05$ ; б)  $\alpha = 0,1$ ? Контрольована ознака має гауссівський розподіл.

**5.12.** Очікується, що при додаванні спеціальних речовин жорсткість води зменшиться. За оцінками жорсткості води до і після додавання

спеціальних речовин по 40 і 50 пробам відповідно отримали середнє значення жорсткості (у стандартних одиницях), що рівні 4,0 і 3,8. Дисперсія вимірів в обох випадках припускається рівною 0,25. Чи підтверджують ці результати очікуваний ефект? Прийняти  $\alpha = 0,05$ . Контрольована ознака має гауссівський розподіл.

**5.13.** Два штурмана визначали пеленг маяка за декількома замірами, використовуючи різні пеленгатори. Результати замірів:  $\bar{x} = 70,2^\circ$  при  $n_1 = 4$  і  $\bar{y} = 70,5^\circ$  при  $n_2 = 9$ . Перевірте при  $\alpha = 0,05$  гіпотезу про те, що різні результати викликані тільки випадковими помилками, якщо середні квадратичні відхилення для обох пеленгаторів відомі і дорівнюють  $\sigma_1 = \sigma_2 = 0,5^\circ$ . Контрольована ознака має гауссівський розподіл.

**5.14.** Заводи А і В випускають прилади одного типу. За вибіркою із 50 приладів заводу А встановили середню тривалість роботи приладу (значення вибіркового середнього) 1288 г із середнім квадратичним відхиленням 80 г, а також за вибіркою того ж об'єму з заводу В – 1208 г із середнім квадратичним відхиленням 94 г. На рівні значущості  $\alpha = 0,05$  перевірити гіпотезу про те, що середній термін роботи приладів на обох заводах однаковий. Вважати, що тривалість роботи одного приладу розподілена приблизно за гауссівським законом.

**5.15.** При обробці втулок на верстаті-автоматі йдуть спостереження за режимом його роботи. Для перевірки стабільності роботи верстата через визначені проміжки часу вивчають вибірку обсягу  $n = 10$ . За результатами двох вибірок (табл. 5.1) перевірити стабільність роботи верстата. Розподіл контрольованої ознаки припускається гауссівським. Також припускається,

що дисперсії генеральних сукупностей, з яких отримані вибірки, рівні. Рівень значущості  $\alpha = 0,05$ .

Таблиця 5.1

Номер виробу	1	2	3	4	5	6	7	8	9	10
$x_i$	2,060	2,063	2,068	2,060	2,067	2,063	2,059	2,062	2,062	2,060
$y_i$	2,063	2,060	2,057	2,056	2,059	2,058	2,062	2,059	2,059	2,057

**5.16.** Точність налагодження верстата-автомата, що виробляє деякі деталі, характеризується дисперсією довжини деталей. Якщо ця величина буде більшою  $400 \text{ мкм}^2$ , верстат зупиняється для налагодження. Значення виправленої вибіркової дисперсії, що знайдене по 15 випадково відібраним деталям з продукції верстату, виявилось рівним  $680 \text{ мкм}^2$ . Визначте, чи потрібне налагодження, якщо: а)  $\alpha = 0,01$ ; б)  $\alpha = 0,1$ . Контрольована ознака має гауссівський розподіл.

**5.17.** При зміні процедури перевірки коефіцієнта тертя встановлено, що дисперсія результатів вимірів цього коефіцієнта складає  $0,1$ . Значення виправленої вибіркової дисперсії, обраховане за результатами 26 вимірів коефіцієнта тертя, виявилось рівним  $0,2$ . На рівні значущості  $\alpha = 0,1$  перевірити гіпотезу про те, що дисперсія результатів вимірів коефіцієнта тертя рівна  $0,1$ . Припускається, що контрольована ознака має гауссівський розподіл.

**5.18.** На двох токарних верстатах-автоматах виготовляють деталі по одному кресленню. Із продукції першого верстата було відібрано  $n_1 = 9$  деталей, а з продукції другого  $n_2 = 11$  деталей. Оцінки виправлених вибіркової дисперсії контрольного розміру, визначені за цими вибірками,

рівні  $(\tilde{D}\xi)_{\text{зн}} = 5,9 \text{ мкм}^2$  і  $(\tilde{D}\eta)_{\text{зн}} = 23,3 \text{ мкм}^2$  відповідно. Перевірте гіпотезу про рівність дисперсійна рівні значущості  $\alpha = 0,05$ . Припускається, що контрольована ознака має гауссівський розподіл.

**5.19.** Тиск у камері контролюється за двома манометрами. Для порівняння точності цих приладів одночасно контролюють їх показники. За результатами замірів значення оцінок (в одиницях шкали приладів) виявились наступними:  $\bar{x} = 1573$ ,  $\bar{y} = 1671$ ,  $(\tilde{D}\xi)_{\text{зн}} = 0,72$ ,  $(\tilde{D}\eta)_{\text{зн}} = 0,15$ . При  $\alpha = 0,1$  перевірити гіпотезу про рівність дисперсій, якщо контрольна ознака має гауссівський розподіл.

**5.20.** На двох верстатах А і В виготовляють одну і ту саму продукцію, що контролюється по внутрішньому діаметру виробу. Із продукції верстата А була взята вибірка із 16 виробів, а із продукції верстата В – вибірка із 25 виробів і отримані значення  $\bar{x}_A = 36,5 \text{ мм}$ ,  $(\tilde{D}\xi)_{\text{зн}} = 1,21 \text{ мм}^2$ ,  $\bar{x}_B = 36,8 \text{ мм}$ ,  $(\tilde{D}\eta)_{\text{зн}} = 1,44 \text{ мм}^2$ . Перевірте гіпотезу про рівність математичних сподівань контрольованих розмірів в продукції обох верстатів якщо: а)  $\alpha = 0,05$ , б)  $\alpha = 0,1$ . Припускається, що розподіл контрольованих розмірів гауссівський.

**5.21.** Порівнюються характеристики міцності сталі марок А і В. Для цього випробувані на границю міцності 145 зразків марки А і 200 зразків марки В. У результаті отримали  $\bar{x}_1 = 31,40$ ,  $(\tilde{D}\xi)_{\text{зн}} = 3,36$ ,  $\bar{x}_2 = 28,84$ ,  $(\tilde{D}\eta)_{\text{зн}} = 3,51$ . Чи можна на рівні значущості  $\alpha = 0,1$  вважати, що сталь двох марок мають різні характеристики міцності? Контрольована ознака має гауссівський розподіл.

**5.22.** Партія виробів приймається, якщо ймовірність того, що виріб буде бракованим не перевищить 0,02. Серед випадково відібраних 1000 виробів виявилось 40 бракованих. Чи можна на рівні значущості  $\alpha = 0,01$  прийняти цю партію?

**5.23.** При 600 підкиданнях гральної кістки шість очок з'явилося 75 разів. На рівні значущості  $\alpha = 0,05$  чи можна вважати, що кістка симетрична? Як ще можна розв'язати цю задачу?

**5.24.** В 105 експериментах подія А відбулася 42 рази. Повторна серія дослідів складалася з 195 експериментів, в яких подія А відбулася 65 разів. Чи можна вважати, що ймовірність появи події А в обох серіях експериментів одна й та ж сама? Гіпотезу перевірити на рівні значущості  $\alpha = 0,01$ .

**5.25.** Для вивчення ефективності профілактичних ліків проти алергії проводилося обстеження двох груп людей, схильних до цієї хвороби. Результати дослідження наведені у таблиці

Приймали ліки		Не приймали ліки	
захворіли	не захворіли	захворіли	не захворіли
3	172	32	168

Чи можна на рівні значущості  $\alpha = 0,05$  сказати, що ці результати доводять ефективність профілактичних ліків проти алергії?

## Відповіді

### До глави 1

**1.1.** Так. **1.2.**  $\theta_1^*$ ,  $\theta_2^*$ ,  $\theta_3^*$ ,  $\theta_4^*$  - дивись приклад 1.1.1,  $\theta_5^*$  є асимптотично незміщеною та конзистентною оцінкою  $b-a$ ,  $\theta_6^*$ ,  $\theta_7^*$ ,  $\theta_8^*$  є незміщеними та конзистентними оцінками відповідно  $\frac{a+b}{2}$ ,  $a$ ,  $b$ . **1.3.** Незміщена та конзистентна оцінка. **1.4.** Так. **1.5.** Найбільшого значення  $L(\theta)$  набуває у точках  $\theta_\mu^* = \mu(\max \xi_i - 1) + (1 - \mu)(\min \xi_i + 1)$ ,  $\mu \in [0,1]$ . Незміщеність та конзистентність цієї оцінки дивимось на прикладі 1.1.1. **1.6.**  $\theta_1^*$ ,  $\theta_2^*$  асимптотично незміщені та конзистентні оцінки параметрів  $\theta$  та  $\alpha$  відповідно. **1.7.** Так. **1.8.**  $\theta^*$  є незміщеною та конзистентною оцінкою параметра  $\theta$ . **1.9.**  $a^*$  є незміщеною та конзистентною оцінкою параметра  $a$ ,  $(\sigma^2)^*$  є асимптотично незміщеною та конзистентною оцінкою параметра  $\sigma^2$ . **1.10.**  $\theta^* = \bar{\xi}$  є незміщеною та конзистентною оцінкою параметра  $\theta$ ,  $(\sigma^2)^* = \bar{\xi}^2 - (\bar{\xi})^2$  є асимптотично незміщеною та конзистентною оцінкою параметра  $\sigma^2$ . **1.11.** Так. **1.12.**  $a^*$  є незміщеною та конзистентною оцінкою параметра  $a$ ,  $\sigma^{2*}$  є асимптотично незміщеною та конзистентною оцінкою параметра  $\sigma^2$  де  $a^* = \bar{\xi}$ ,  $(\sigma^2)^* = \bar{\xi}^2 - (\bar{\xi})^2$ . **1.13.** Так. **1.14.** Так. **1.15.** Так. **1.16.**  $a_2^*$  незміщеною оцінкою параметра  $\alpha$ , але не є конзистентною оцінкою. **1.17.** Так. **1.18.**  $\theta_1^*$  є незміщеною та конзистентною оцінкою  $\theta + b$ ,  $\theta_2^*$  та  $\theta_3^*$  є асимптотично незміщені та конзистентні оцінки параметрів  $b$  та  $\theta$ . **1.19.** Так. **1.20.**  $\theta_3^*$  є незміщеною та конзистентною оцінкою параметра  $\frac{a_0+b}{2}$ ,  $\theta_2^*$  та  $\theta_4^*$  асимптотично незміщені та конзистентні оцінки параметра  $b$ . **1.21.** Так. **1.22.** Дивись приклад 1.1.1 та задачу 1.2.

**1.23.** Так. **1.24.** Дивись приклад 1.1.1 та задачу 1.2. **1.25.** Ні. **1.26.** Так. **1.27.** Так. **1.28.** Так.

## До глави 2

**2.1.** а)  $\bar{x} = 16,46$ ;  $(Me^*\xi)_{\text{ЗН}} = (Mo^*\xi)_{\text{ЗН}}=16$ ;  $(\tilde{D}\xi)_{\text{ЗН}} = 4,92$ ; б)  $\bar{x} = 11,32$ ;  $(Me^*\xi)_{\text{ЗН}} = 10,75$ ;  $(Mo^*\xi)_{\text{ЗН}} = 11$ ;  $(\tilde{D}\xi)_{\text{ЗН}} = 14,7$ . **2.2.**  $\bar{x} = 3,2$ ;  $(Me^*\xi)_{\text{ЗН}} = (Mo^*\xi)_{\text{ЗН}}=3$ ;  $(\tilde{D}\xi)_{\text{ЗН}} = 5,96$ . **2.3.** а) збільшиться (зменшиться) на  $d$ ; б) )

збільшиться (зменшиться) в  $k$  разів. **2.4.** а) незміниться; б) збільшиться (зменшиться) в  $k^2$  разів. **2.5.**  $p^* = \frac{1}{\bar{\xi}+1}$ . **2.7.**  $(D\xi)^* = (2\theta^2)^* = \bar{\xi}^2$  – незміщена

та консистентна. **2.8.**  $\theta^* = \bar{\xi}$ ,  $(\sigma^2)^* = D^*\xi$ . **2.9.**  $\theta^* = \frac{(\bar{\xi}^2)}{2}$  – незміщена,

консистентна та ефективна. **2.10.**  $\lambda^* = \frac{\bar{\xi}}{D^*\xi}$ ,  $m^* = \frac{(\bar{\xi})^2}{D^*\xi}$ . **2.12.** Дивись

приклад 2.2.1. **2.13.**  $\theta^* = \bar{\xi}$  – незміщена, консистентна та ефективна.

**2.14.**  $\theta^* = \bar{\xi}$ ,  $(\sigma^2)^* = D^*\xi$ . **2.15.**  $p^* = \frac{\bar{\xi}-D^*\xi}{\bar{\xi}}$ ,  $m^* = \frac{(\bar{\xi})^2}{\bar{\xi}-D^*\xi}$ . **2.16.**  $\theta^* = 1 + \frac{\sqrt{\bar{\xi}^2}}{\sigma^*}$ ,

$\lambda^* = \frac{\bar{\xi}\sqrt{\bar{\xi}^2}}{\sigma^* + \sqrt{\bar{\xi}^2}}$ . **2.17.**  $\theta^* = \frac{\bar{\xi}}{r}$  – незміщена, консистентна та ефективна.

**2.18.**  $p^* = \frac{\bar{\xi}}{m}$  – незміщена, консистентна та ефективна. **2.19.**  $\lambda^* = \bar{\xi}$  –

незміщена, консистентна та ефективна. **2.20.**  $p^* = \frac{k}{n}$  ( $k$  – кількість успішних дослідів) – незміщена, консистентна та ефективна.

**2.21.**  $p^*(\vec{\xi}) = \frac{1}{\bar{\xi}}$ . **2.22.**  $\beta^*(\vec{\xi}) = \frac{\bar{\xi}}{\alpha+1}$ . **2.23.**  $\alpha^*(\vec{\xi}) = \frac{k}{\bar{\xi}}$ .

**2.24.**  $\theta^*(\vec{\xi}) = \max_{i=\overline{1,n}} |\xi_i|$ .

### До глави 3

**3.1.** а)  $(\bar{\xi} - 1,55; \bar{\xi} + 1,55)$ ; б)  $(\bar{\xi} - 1,09; \bar{\xi} + 1,09)$ ; в)  $(\bar{\xi} - 0,69; \bar{\xi} + 0,69)$ ;  
**3.2.** На літаку повинно бути не менше двох висотомірів. **3.3.** (5,25; 5,75).  
**3.4.** а) 0,68; б)  $n \geq 385$ . **3.5.**  $\bar{x} = 4,783 \cdot 10^{-10}$ ;  $(4,761 \cdot 10^{-10};$   
 $4,805 \cdot 10^{-10})$ . **3.6.** а) (2991,2; 3008,8); б) 0,93. **3.7.** 0,55. **3.8.** (27,69; 30,74).  
**3.9.** (90,32; 480,83). **3.10.** а) (6,69; 12,7); б) 0,76. **3.11.** не менше 15 вимірів.  
**3.12.** (4,12; 15,88). **3.13.** (0,266N; 0,454N). **3.14.** а) (0,051; 0,149); б)  $n \geq 324$ .  
**3.15.** а) (0,39; 0,41); б)  $n \geq 16231$ . **3.16. Вказівка:** в якості вихідної  
розгляньте статистику  $T = X_{(1)} - \theta$ , що має функцію розподілу

$$F(x) = \begin{cases} 1 - e^{-nx}, & \text{якщо } x \geq 0; \\ 0, & \text{якщо } x < 0. \end{cases}$$
$$\left( X_{(1)} + \frac{\ln \beta}{n}; X_{(1)} - \frac{\ln(1-\alpha)}{n} \right),$$

де  $\alpha > 0$  і  $\beta > 0$  пов'язані рівністю  $1 - \alpha - \beta = \gamma$ .

**3.17.** (-2, 052; 2,852). **3.18.** а) (2991,235; 3008,765); б) 0,93. **3.19.** а) (15,32;  
24,68); б) 0,71; в) 201.

### До глави 4

**4.1.** а) Гіпотеза приймається. **4.2.** в) Гіпотеза приймається. **4.3.** Гіпотеза  
приймається. **4.4.** Так. **4.5.** Так. **4.6.** Так. **4.7.** Гіпотеза незалежності  
відхиляється. **4.14.** Гіпотеза відхиляється. **4.15.** Гіпотеза незалежності  
відхиляється. **4.16.** Гіпотеза незалежності приймається.

### До глави 5

**5.1.** а) кількість бракованих виробів, область значень  $\{0, 1, \dots, 10\}$ , критична  
область  $\{2, 3, \dots, 10\}$ ; б) біноміальний з параметрами 10 та  $p$ ;

в)  $H_0: p = 0,05$ , вірне твердження постачальника,  $H_1: p = 0,1$ , вірне твердження покупця; г) помилка першого роду: партія прийнята на умовах покупця, хоча вірне твердження постачальника; помилка другого роду: партія прийнята на умовах постачальника, хоча вірне твердження покупця,  $\alpha \approx 0,086$ ,  $\beta \approx 0,736$ . **5.2.** а) критична область  $\{0,1,2,3,4\}$ ; б) біноміальний з параметрами 5 та  $p$ ; в)  $H_0: p = 0,99$ , пристрій працює правильно;  $H_1: p = 0,4$ , пристрій працює неправильно, помилка другого роду – прийняття неправильно працюючого пристрою; г)  $\alpha \approx 0,05$ ,  $\beta \approx 0,01$ . **5.3.** а) так; б)  $(40,32, +\infty)$ . **5.4.** а) так; б) ні. **5.5.** а) ні; б) ні. **5.6.** а) гіпотеза приймається; б) гіпотеза приймається.

**5.7.** Значення потужності **5.8.** ні, не можна. **5.9.** а) гіпотеза критерію наведені у таблиці відхиляється; б) гіпотеза приймається.

$d$	$n=100$	$n=16$
10,1	0,001	0,107
10,2	0,017	0,201
10,3	0,134	0,330

**5.10.**  $1 - \beta = 0,323$ ;  $n > 294$ . **5.11.** а) так; б) так. **5.13.** гіпотеза приймається. **5.14.** гіпотеза відхиляється. **5.15.** гіпотезу про стабільність роботи станка слід відхилити.

**5.16.** а) ні; б) так. **5.17.** гіпотеза відхиляється. **5.18.** гіпотеза відхиляється. **5.19.** гіпотеза приймається. **5.20.** а) гіпотеза приймається; б) гіпотеза приймається. **5.21.** так. **5.22.** ні. **5.23.** ні. Задачу також можна розв'язати за критерієм  $\chi^2$ . **5.24.** так. **5.25.** так.

## Список літератури

1. Гихман И.И. Теория вероятностей и математическая статистика / Гихман И.И., Скороход А.В., Ядренко М.И. – К.: Вища школа, 1979. – 408 с.
2. Горяинов В.Б. Математическая статистика / Горяинов В.Б., Павлов И.В., Цветкова Г.М. – М.: Изд-во МГТУ им. Н.Э.Баумана, 2001. 423 с.
3. Крамер Г. Математические методы статистики / Крамер Г. – М.: Мир, 1975. 648 с
4. Кремер Н.Ш. Теория вероятностей и математическая статистика / Кремер Н.Ш. – М.: ЮНИТИ, 2007. 573 с.
5. Сборник задач по математике для вузов: Специальные курсы / Под ред. А.В. Ефимова. – М.: Наука, 1984.– 606 с.
6. Турчин В.М. Математична статистика. Навч. посіб. / Турчин В.М. К.: Академія, 1999. 240 с
7. Черняк І.О. Теорія ймовірностей та математична статистика. Збірник задач. / Черняк І.О., Обушна О.М., Ставицький А.В. К.: Знання, 2002. 200 с.

## Додатки. Таблиці математичної статистики

### Додаток 1

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

<i>x</i>	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
<b>0,00</b>	0,00000	0,00399	0,00798	0,01197	0,01596	0,01994	0,02392	0,02791	0,03188	0,03586
<b>0,10</b>	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06750	0,07143	0,07535
<b>0,20</b>	0,07926	0,08317	0,08707	0,09096	0,09484	0,09871	0,10257	0,10642	0,11026	0,11409
<b>0,30</b>	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
<b>0,40</b>	0,15542	0,15910	0,16276	0,16640	0,17003	0,17365	0,17724	0,18082	0,18439	0,18794
<b>0,50</b>	0,19146	0,19498	0,19847	0,20195	0,20540	0,20884	0,21226	0,21566	0,21905	0,22241
<b>0,60</b>	0,22575	0,22907	0,23237	0,23566	0,23892	0,24216	0,24538	0,24857	0,25175	0,25491
<b>0,70</b>	0,25804	0,26115	0,26424	0,26731	0,27035	0,27338	0,27638	0,27935	0,28231	0,28524
<b>0,80</b>	0,28815	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
<b>0,90</b>	0,31594	0,31859	0,32122	0,32382	0,32639	0,32895	0,33147	0,33398	0,33646	0,33892
<b>1,00</b>	0,34135	0,34375	0,34614	0,34850	0,35083	0,35314	0,35543	0,35769	0,35993	0,36215
<b>1,10</b>	0,36434	0,36650	0,36865	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
<b>1,20</b>	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40148
<b>1,30</b>	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41309	0,41466	0,41621	0,41774
<b>1,40</b>	0,41925	0,42073	0,42220	0,42364	0,42507	0,42647	0,42786	0,42922	0,43057	0,43189
<b>1,50</b>	0,43320	0,43348	0,43575	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44409
<b>1,60</b>	0,44520	0,44630	0,44739	0,44845	0,44950	0,45053	0,45155	0,45254	0,45352	0,45449
<b>1,70</b>	0,45544	0,45637	0,45729	0,45819	0,45907	0,45994	0,46080	0,46164	0,46246	0,46328
<b>1,80</b>	0,46407	0,46485	0,46562	0,46638	0,46712	0,46785	0,46856	0,46926	0,46995	0,47062
<b>1,90</b>	0,47129	0,47194	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47671
<b>2,00</b>	0,47725	0,47779	0,47831	0,47882	0,47933	0,47982	0,48030	0,48078	0,48124	0,48169
<b>2,10</b>	0,48214	0,48257	0,48300	0,48342	0,48383	0,48422	0,48462	0,48500	0,48537	0,48574
<b>2,20</b>	0,48610	0,48645	0,48679	0,48713	0,48746	0,48778	0,48809	0,48840	0,48870	0,48899
<b>2,30</b>	0,48928	0,48956	0,48983	0,49010	0,49036	0,49062	0,49086	0,49111	0,49135	0,49158
<b>2,40</b>	0,49180	0,49203	0,49224	0,49245	0,49266	0,49286	0,49306	0,49325	0,49343	0,49362
<b>2,50</b>	0,49379	0,49397	0,49413	0,49430	0,49446	0,49462	0,49477	0,49492	0,49506	0,49520
<b>2,60</b>	0,49534	0,49548	0,49561	0,49573	0,49586	0,49598	0,49610	0,49621	0,49632	0,49643
<b>2,70</b>	0,49654	0,49664	0,49674	0,49684	0,49693	0,49703	0,49711	0,49720	0,49728	0,49737
<b>2,80</b>	0,49745	0,49753	0,49760	0,49768	0,49775	0,49782	0,49788	0,49795	0,49801	0,49808

$x$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
<b>2,90</b>	0,49814	0,49820	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861
<b>3,00</b>	0,49865	0,49870	0,49874	0,49878	0,49882	0,49886	0,49890	0,49893	0,49897	0,49900
<b>3,10</b>	0,49903	0,49907	0,49910	0,49913	0,49916	0,49919	0,49921	0,49924	0,49927	0,49929
<b>3,20</b>	0,49932	0,49934	0,49936	0,49938	0,49940	0,49943	0,49945	0,49946	0,49948	0,49950
<b>3,30</b>	0,49952	0,49954	0,49955	0,49957	0,49958	0,49960	0,49961	0,49963	0,49964	0,49965
<b>3,40</b>	0,49967	0,49968	0,49969	0,49970	0,49971	0,49522	0,49973	0,49974	0,49975	0,49976
<b>3,50</b>	0,49977	0,49978	0,49979	0,49979	0,49980	0,49981	0,49982	0,49982	0,49983	0,49984
<b>3,60</b>	0,49984	0,49985	0,49986	0,49986	0,49987	0,49988	0,49988	0,49988	0,49989	0,49989
<b>3,70</b>	0,49989	0,49990	0,49990	0,49991	0,49991	0,49991	0,49992	0,49992	0,49992	0,49993
<b>3,80</b>	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995	0,49995
<b>3,90</b>	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997	0,49997
<b>4,00</b>	0,49997	0,49997	0,49997	0,49997	0,49998	0,49998	0,49998	0,49998	0,49998	0,49998
<b>4,10</b>	0,49998	0,49998	0,49998	0,49998	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999
<b>4,20</b>	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999	0,49999

**Додаток 2**

**$\chi^2$ -розподіл.** Значення функції  $\chi_{\alpha; n}^2$  визначається з рівняння  $\int_{\chi_{\alpha; n}^2}^{+\infty} f(x)dx = \alpha$ , де  $f(x)$  – щільність  $\chi_n^2$ -розподілу (розподілу Пірсона з  $n$  ступенями вільності).

$$f(x) = \begin{cases} \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}}, & \text{якщо } x > 0; \\ 0, & \text{якщо } x \leq 0. \end{cases}$$

<b><math>n</math></b>	<b>Значення <math>\alpha</math></b>							
	<b>0,990</b>	<b>0,975</b>	<b>0,950</b>	<b>0,900</b>	<b>0,100</b>	<b>0,050</b>	<b>0,025</b>	<b>0,010</b>
<b>1</b>	0,00	0,00	0,00	0,02	2,71	3,84	5,02	6,64
<b>2</b>	0,02	0,05	0,10	0,21	4,61	5,99	7,38	9,21
<b>3</b>	0,12	0,22	0,35	0,58	6,23	7,82	9,35	11,34
<b>4</b>	0,30	0,48	0,71	1,06	7,78	9,48	11,14	13,28
<b>5</b>	0,55	0,83	1,14	1,61	9,24	11,07	12,83	15,09
<b>6</b>	0,87	1,24	1,64	2,20	10,64	12,59	14,45	16,81
<b>7</b>	1,24	1,69	2,17	2,83	12,02	14,07	16,01	18,48
<b>8</b>	1,65	2,18	2,73	3,49	13,36	15,51	17,53	20,09
<b>9</b>	2,09	2,70	3,32	4,17	14,68	16,92	19,02	21,67
<b>10</b>	2,56	3,25	3,94	4,86	15,99	18,31	20,48	23,21
<b>11</b>	3,05	3,82	4,58	5,58	17,28	19,68	21,92	24,72
<b>12</b>	3,57	4,40	5,23	6,30	18,55	21,03	23,34	26,22
<b>13</b>	4,11	5,01	5,89	7,04	19,81	22,36	24,74	27,69
<b>14</b>	4,66	5,63	6,57	7,79	21,06	23,68	26,12	29,14
<b>15</b>	5,23	6,26	7,26	8,55	22,31	25,00	27,49	30,58
<b>16</b>	5,81	6,91	7,96	9,31	23,54	26,30	28,85	32,00
<b>17</b>	6,41	7,56	8,67	10,09	24,77	27,59	30,19	33,41
<b>18</b>	7,02	8,23	9,39	10,86	25,99	28,87	31,53	34,81
<b>19</b>	7,63	8,91	10,12	11,65	27,20	30,14	32,85	36,19
<b>20</b>	8,26	9,59	10,85	12,44	28,41	31,41	34,17	37,57
<b>21</b>	8,90	10,28	11,59	13,24	29,62	32,67	35,48	38,93
<b>22</b>	9,54	10,98	12,34	14,04	30,81	33,92	36,78	40,29

<b><i>n</i></b>	<b>Значення <math>\alpha</math></b>							
	<b>0,990</b>	<b>0,975</b>	<b>0,950</b>	<b>0,900</b>	<b>0,100</b>	<b>0,050</b>	<b>0,025</b>	<b>0,010</b>
<b>23</b>	10,20	11,69	13,09	14,85	32,01	35,17	38,08	41,64
<b>24</b>	10,86	12,40	13,85	15,66	33,20	36,42	39,36	42,92
<b>25</b>	11,52	13,12	14,61	16,47	34,38	37,65	40,65	44,31
<b>26</b>	12,20	13,84	15,38	17,29	35,56	38,89	41,92	45,64
<b>27</b>	12,88	14,57	16,15	18,11	36,74	40,11	43,19	46,96
<b>28</b>	13,56	15,31	16,93	18,94	37,92	41,34	44,46	48,26
<b>29</b>	14,26	16,05	17,71	19,77	39,09	42,56	45,72	49,59
<b>30</b>	14,95	16,79	18,49	20,60	40,26	43,77	46,98	50,89
<b>40</b>	22,16	24,43	26,51	29,05	51,80	55,76	59,34	63,69
<b>50</b>	29,71	32,36	34,76	37,69	63,17	67,50	71,42	76,15
<b>60</b>	37,48	40,48	43,19	46,46	74,40	79,08	83,30	88,38
<b>70</b>	45,44	48,76	51,74	55,33	85,53	90,53	95,02	100,4
<b>80</b>	53,54	57,15	60,39	64,28	96,58	101,9	106,6	112,3
<b>90</b>	61,75	65,65	69,13	73,29	107,6	113,1	118,1	124,1
<b>100</b>	70,06	74,22	77,93	82,36	118,5	124,3	129,6	135,8

**Розподіл Стьюдента.** Значення функції  $t_{\alpha;n}$  визначається з рівняння  $\int_{t_{\alpha;n}}^{+\infty} f(x)dx = \alpha$ , де  $f(x)$  – щільність розподілу Стьюдента з  $n$  ступенями вільності.

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

<b><i>n</i></b>	<b>Значення <math>\alpha</math></b>			
	<b>0,050</b>	<b>0,025</b>	<b>0,010</b>	<b>0,005</b>
<b>1</b>	6,314	12,706	31,821	63,657
<b>2</b>	2,920	4,303	6,965	9,925
<b>3</b>	2,353	3,182	4,541	5,841
<b>4</b>	2,132	2,776	3,747	4,604
<b>5</b>	2,015	2,571	3,365	4,032
<b>6</b>	1,943	2,447	3,143	3,707
<b>7</b>	1,895	2,365	2,998	3,499
<b>8</b>	1,860	2,306	2,896	3,355
<b>9</b>	1,833	2,262	2,821	3,250
<b>10</b>	1,812	2,228	2,764	3,169
<b>11</b>	1,796	2,201	2,718	3,106
<b>12</b>	1,782	2,179	2,681	3,055
<b>13</b>	1,771	2,160	2,650	3,012
<b>14</b>	1,761	2,145	2,624	2,977
<b>15</b>	1,753	2,131	2,602	2,947
<b>16</b>	1,746	2,120	2,583	2,921
<b>17</b>	1,740	2,110	2,567	2,898
<b>18</b>	1,734	2,101	2,552	2,878
<b>19</b>	1,729	2,093	2,539	2,861
<b>20</b>	1,725	2,086	2,528	2,845
<b>21</b>	1,721	2,080	2,518	2,831
<b>22</b>	1,717	2,074	2,508	2,819
<b>23</b>	1,714	2,069	2,500	2,807
<b>24</b>	1,711	2,064	2,492	2,797

<i>n</i>	Значення $\alpha$			
	0,050	0,025	0,010	0,005
25	1,708	2,060	2,485	2,787
26	1,706	2,056	2,479	2,779
27	1,703	2,052	2,473	2,771
28	1,701	2,048	2,467	2,763
29	1,699	2,045	2,462	2,756
30	1,697	2,042	2,457	2,750
40	1,684	2,021	2,423	2,704
50	1,676	2,009	2,403	2,678
60	1,671	2,000	2,390	2,660
70	1,667	1,994	2,381	2,648
80	1,664	1,990	2,374	2,639
90	1,662	1,987	2,368	2,632
100	1,660	1,984	2,364	2,626
120	1,658	1,980	2,358	2,617
$\infty$	1,645	1,960	2,326	2,576

#### Додаток 4

**Розподіл Фішера.** Значення функції  $F_{\alpha; n; m}$  визначається з рівняння  $\int_{F_{\alpha; n; m}}^{+\infty} f(x)dx = \alpha$ , де  $f(x)$  – щільність розподілу Фішера з  $n, m$  ступенями вільності.

$$f(x) = \begin{cases} \frac{\Gamma\left(\frac{n+m}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{m}{2}\right)} \left(\frac{n}{m}\right)^{\frac{n}{2}} \frac{x^{\frac{n}{2}-1}}{\left(\frac{n}{m}x+1\right)^{\frac{n+m}{2}}}, & \text{якщо } x > 0; \\ 0, & \text{якщо } x \leq 0. \end{cases}$$

Значення функції  $F_{0,05; n; m}$

$m$	$n$ (число ступенів вільності чисельника)									
	1	2	3	4	5	6	7	8	9	10
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93

Значення функції  $F_{0,05; n; m}$

<b><i>m</i></b>	<b><i>n</i> (число ступенів вільності чисельника)</b>									
	<b>12</b>	<b>14</b>	<b>16</b>	<b>18</b>	<b>20</b>	<b>30</b>	<b>40</b>	<b>50</b>	<b>60</b>	<b>100</b>
<b>3</b>	8,74	8,71	8,69	8,67	8,66	8,62	8,59	8,58	8,57	8,55
<b>4</b>	5,91	5,87	5,84	5,82	5,80	5,75	5,72	5,70	5,69	5,66
<b>5</b>	4,68	4,64	4,60	4,58	4,56	4,50	4,46	4,44	4,43	4,41
<b>6</b>	4,00	3,96	3,92	3,90	3,87	3,81	3,77	3,75	3,74	3,71
<b>7</b>	3,57	3,53	3,49	3,47	3,44	3,38	3,34	3,32	3,30	3,27
<b>8</b>	3,28	3,24	3,20	3,17	3,15	3,08	3,04	3,02	3,01	2,97
<b>9</b>	3,07	3,03	2,99	2,96	2,94	2,86	2,83	2,80	2,79	2,76
<b>10</b>	2,91	2,86	2,83	2,80	2,77	2,70	2,66	2,64	2,62	2,59
<b>11</b>	2,79	2,74	2,70	2,67	2,65	2,57	2,53	2,51	2,49	2,46
<b>12</b>	2,69	2,64	2,60	2,57	2,54	2,47	2,43	2,40	2,38	2,35
<b>13</b>	2,60	2,55	2,51	2,48	2,46	2,38	2,34	2,31	2,30	2,26
<b>14</b>	2,53	2,48	2,44	2,41	2,39	2,31	2,27	2,24	2,22	2,19
<b>15</b>	2,48	2,42	2,38	2,35	2,33	2,25	2,20	2,18	2,16	2,12
<b>16</b>	2,42	2,37	2,33	2,30	2,28	2,19	2,15	2,12	2,11	2,07
<b>17</b>	2,38	2,33	2,29	2,26	2,23	2,15	2,10	2,08	2,06	2,02
<b>18</b>	2,34	2,29	2,25	2,22	2,19	2,11	2,06	2,04	2,02	1,98
<b>19</b>	2,31	2,26	2,21	2,18	2,16	2,07	2,03	2,00	1,98	1,94
<b>20</b>	2,28	2,22	2,18	2,15	2,12	2,04	1,99	1,97	1,95	1,91
<b>22</b>	2,23	2,17	2,13	2,10	2,07	1,98	1,94	1,91	1,89	1,85
<b>24</b>	2,18	2,13	2,09	2,05	2,03	1,94	1,89	1,86	1,84	1,80
<b>26</b>	2,15	2,09	2,05	2,02	1,99	1,90	1,85	1,82	1,80	1,76
<b>28</b>	2,12	2,06	2,02	1,99	1,96	1,87	1,82	1,79	1,77	1,73
<b>30</b>	2,09	2,04	1,99	1,96	1,93	1,84	1,79	1,76	1,74	1,70
<b>40</b>	2,00	1,95	1,90	1,87	1,84	1,74	1,69	1,66	1,64	1,59
<b>50</b>	1,95	1,89	1,85	1,81	1,78	1,69	1,63	1,60	1,58	1,52
<b>60</b>	1,92	1,86	1,82	1,78	1,75	1,65	1,59	1,56	1,53	1,48
<b>100</b>	1,85	1,79	1,75	1,71	1,68	1,57	1,52	1,48	1,45	1,39