

Домашня Робота з Математичної Статистики #2

Захаров Дмитро

8 вересня, 2024

Зміст

1	Вправи з лекції	2
1.1	Вправа 1. Характеристики розподілу χ_n^2	2
1.2	Вправа 2. Вибіркова дисперсія	3
1.3	Вправа 3. Незміщена оцінка дисперсії	4
1.4	Вправа 4. Слушність незміщеної оцінки дисперсії	5
1.5	Вправа 5. Вибірковий центральний момент порядку 1	5
1.6	Вправа 6. Груповані дані	6
2	Вправи з практики	7
2.1	Вправа 1. Укладені контракти	7
2.2	Вправа 2. Час на виготовлення деталі	9

1 Вправи з лекції

1.1 Вправа 1. Характеристики розподілу χ_n^2

Умова Задачі 1.1. Нехай $\xi \sim \chi_n^2$. Довести, що $\mathbb{E}[\xi] = n$, $\text{Var}[\xi] = 2n$.

Розв'язання. За означенням величина ξ дорівнює:

$$\xi = \sum_{j=1}^n \xi_j^2, \quad (\xi_1, \dots, \xi_n) \sim \mathcal{N}(\mathbf{0}_n, E_{n \times n}).$$

Знаходимо математичне сподівання:

$$\mathbb{E}[\xi] = \mathbb{E}\left[\sum_{j=1}^n \xi_j^2\right] = \sum_{j=1}^n \mathbb{E}[\xi_j^2]$$

Тепер треба знайти $\mathbb{E}[\xi_j^2]$. Можна зробити це означенням, шукаючи $\int_{\mathbb{R}} z^2 f_{\xi_j}(z) dz$, де $f_{\xi_j}(z) = \exp(-z^2/2)/\sqrt{2\pi}$ — щільність розподілу ξ_j . Проте, легше діяти так: ми знаємо, що $\text{Var}[\xi_j] = 1$, а з іншого боку $\text{Var}[\xi_j] = \mathbb{E}[\xi_j^2] - \mathbb{E}[\xi_j]^2$. Математичне сподівання ξ_j дорівнює нулю, тому $\mathbb{E}[\xi_j^2] = 1$. Отже,

$$\mathbb{E}[\xi] = \sum_{j=1}^n \mathbb{E}[\xi_j^2] = \sum_{j=1}^n 1 = n.$$

Тепер знаходимо дисперсію:

$$\text{Var}[\xi] = \text{Var}\left[\sum_{j=1}^n \xi_j^2\right] = \sum_{j=1}^n \text{Var}[\xi_j^2] = \sum_{j=1}^n (\mathbb{E}[\xi_j^4] - \mathbb{E}[\xi_j^2]^2) = \sum_{j=1}^n \mathbb{E}[\xi_j^4] - n$$

Тут ми скористалися тим, що ξ_j^2 — незалежні величини, тому дисперсія суми дорівнює сумі дисперсій. Нарешті, потрібно знайти $\mathbb{E}[\xi_j^4]$. Тут я лише бачу спосіб зробити це за означенням, але одразу розглянемо $\mathbb{E}[\xi_j^{2n}]$:

$$\mathbb{E}[\xi_j^{2n}] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z^{2n} \exp\left\{-\frac{z^2}{2}\right\} dz$$

Будемо інтегрувати частинами. Нехай $du := z^{2n} dz$, $v = \exp\left\{-\frac{z^2}{2}\right\}$. Тоді:

$$\mathbb{E}[\xi_j^{2n}] = \frac{1}{\sqrt{2\pi}} \left(\frac{z^{2n+1}}{2n+1} \exp\left\{-\frac{z^2}{2}\right\} \Big|_{z \rightarrow -\infty}^{z \rightarrow +\infty} + \int_{\mathbb{R}} \frac{z^{2n+1}}{2n+1} \cdot z \exp\left\{-\frac{z^2}{2}\right\} dz \right)$$

Звичайно, що границя всередині дужок зануляється, оскільки експонента спадає швидше за будь-яку степінь полінома. Тому:

$$\mathbb{E}[\xi_j^{2n}] = \frac{1}{(2n+1)\sqrt{2\pi}} \int_{\mathbb{R}} z^{2n+2} \exp\left\{-\frac{z^2}{2}\right\} dz = \frac{1}{2n+1} \cdot \mathbb{E}[\xi_j^{2n+2}]$$

Отже отримали рекурентну формулу $\mathbb{E}[\xi_j^{2n+2}] = (2n+1)\mathbb{E}[\xi_j^{2n}]$. Зокрема, $\mathbb{E}[\xi_j^4] = 3\mathbb{E}[\xi_j^2] = 3$. Тому:

$$\text{Var}[\xi] = \sum_{j=1}^n \mathbb{E}[\xi_j^4] - n = 3n - n = 2n. \quad \blacksquare$$

1.2 Вправа 2. Вибіркова дисперсія

Умова Задачі 1.2. Нехай маємо $x_1, \dots, x_n \sim X$, причому $\mathbb{E}[X] = \mu$, $\text{Var}[X] = \sigma^2$. **Вибірковою дисперсією** називають точкову оцінку:

$$\bar{\sigma}_X = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

- (а) Доведіть наступну формулу для обчислення вибіркової дисперсії, яку зручно використовувати при її підрахунку:

$$\bar{\sigma}_X^2 = \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) - \bar{x}^2.$$

- (б) Доведіть наступне зображення для вибіркової дисперсії:

$$\bar{\sigma}_X^2 = \left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 \right) - (\bar{x} - \mu)^2.$$

Чому цю формулу не можна зазвичай застосовувати на практиці при обчисленні вибіркової дисперсії?

Розв'язання.

Пункт (а). Просто перетворимо вираз з означення для $\bar{\sigma}_X$:

$$\bar{\sigma}_X = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \frac{2\bar{x}}{n} \sum_{i=1}^n x_i + \frac{1}{n} \sum_{i=1}^n \bar{x}^2$$

Помітимо, що $\sum_{i=1}^n x_i = n\bar{x}$, а тому:

$$\bar{\sigma}_X = \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\bar{x}^2 + \bar{x}^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2. \quad \blacksquare$$

Пункт (б). Знову, алгебраїчно маємо:

$$\bar{\sigma}_X^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - (\bar{x} - \mu)^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - 2\mu\bar{x} + \mu^2 - \bar{x}^2 + 2\mu\bar{x} - \mu^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2. \quad \blacksquare$$

1.3 Вправа 3. Незміщена оцінка дисперсії

Умова Задачі 1.3. Незміщеною оцінкою дисперсії називають вираз:

$$\hat{\sigma}_X^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Чому ця оцінка називається незміщеною?

Розв’язання. Незміщеність означає, що математичне сподівання оцінки дорівнює справжньому значенню параметра, тобто

$$\mathbb{E}[\hat{\sigma}_X^2] = \sigma^2.$$

З лекції показано, що $\hat{\sigma}_X^2 = \frac{n}{n-1} \cdot \bar{\sigma}_X^2$, де $\bar{\sigma}_X^2$ — вибіркова дисперсія. Тому:

$$\mathbb{E}[\hat{\sigma}_X^2] = \frac{n}{n-1} \cdot \mathbb{E}[\bar{\sigma}_X^2]$$

На лекції також доведено, що $\mathbb{E}[\bar{\sigma}_X^2] = \frac{n-1}{n} \cdot \sigma^2$. Отже,

$$\mathbb{E}[\hat{\sigma}_X^2] = \frac{n}{n-1} \cdot \frac{n-1}{n} \cdot \sigma^2 = \sigma^2. \quad \blacksquare$$

1.4 Вправа 4. Слушність незміщеної оцінки дисперсії

Умова Задачі 1.4. Довести слушність незміщеної оцінки дисперсії.

Розв'язання. Потрібно довести, за означенням:

$$\forall \varepsilon > 0 : \lim_{n \rightarrow \infty} \Pr[|\hat{\sigma}_{X,n}^2 - \sigma^2| \geq \varepsilon] = 0.$$

З нерівності Маркова маємо:

$$\Pr[|\hat{\sigma}_{X,n}^2 - \sigma^2| \geq \varepsilon] \leq \frac{\mathbb{E}[|\hat{\sigma}_{X,n}^2 - \sigma^2|]}{\varepsilon}.$$

Звідси мені не вдалося строго довести, що $\mathbb{E}[|\hat{\sigma}_{X,n}^2 - \sigma^2|] \xrightarrow{n \rightarrow \infty} 0$.

1.5 Вправа 5. Вибірковий центральний момент порядку 1

Умова Задачі 1.5. Доведіть, що $\bar{\mu}_1 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x}) = 0$.

Розв'язання.

$$\bar{\mu}_1 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x}) = \frac{1}{n} \sum_{k=1}^n x_k - \frac{1}{n} \sum_{k=1}^n \bar{x} = \bar{x} - \bar{x} = 0. \quad \blacksquare$$

1.6 Вправа 6. Груповані дані

Умова Задачі 1.6. Нехай маємо інтервали $[x_1, x_2), [x_2, x_3), \dots, [x_m, x_{m+1})$ з числами даних, що потрапили у відповідний проміжок, n_1, \dots, n_m . Для обчислення точкових оцінок, заміняємо інтервали на середини $z_j := \frac{x_j + x_{j+1}}{2}$. Запишіть формули для обчислення вибіркового середнього та вибіркової дисперсії в термінах цієї таблиці.

Розв'язання. Будемо вважати, що наш розподіл має вигляд:

$$\Pr[\xi = z_j] = \frac{n_j}{\sum_{\ell=1}^m n_\ell},$$

а $\underbrace{z_1, z_1, \dots, z_1}_{n_1}, \underbrace{z_2, \dots, z_2}_{n_2}, \dots, \underbrace{z_m, \dots, z_m}_{n_m}$ — це вибірка з розподілу ξ . Тоді, згідно означенню, вибіркове середнє $\hat{\mu}$ дорівнює:

$$\hat{\mu} = \frac{1}{\sum_{j=1}^m n_j} \cdot \sum_{j=1}^m n_j z_j = \frac{\sum_{j=1}^m n_j z_j}{\sum_{j=1}^m n_j}$$

Тепер знайдемо вибіркиму дисперсію. За означенням, вона дорівнює:

$$\hat{\sigma}^2 = \frac{1}{\sum_{j=1}^m n_j} \cdot \sum_{j=1}^m n_j z_j^2 - \hat{\mu}^2 = \frac{(\sum_{j=1}^m n_j) (\sum_{j=1}^m n_j z_j^2) - (\sum_{j=1}^m n_j z_j)^2}{(\sum_{j=1}^m n_j)^2}$$

2 Вправи з практики

2.1 Вправа 1. Укладені контракти

Умова Задачі 2.1. Провалились спостереження числа укладених контрактів різними фірмами міста, в результаті яких отримано наступні вибіркові дані про числа контрактів фірмами міста протягом місяця:

Число контрактів	20	40	60	80
Число фірм	10	15	25	10

Побудувати полігон частот, вибірккову функцію розподілу та гістограму вибірки. Знайти вибірккове середнє, вибірккову дисперсію та незміщену оцінку дисперсії.

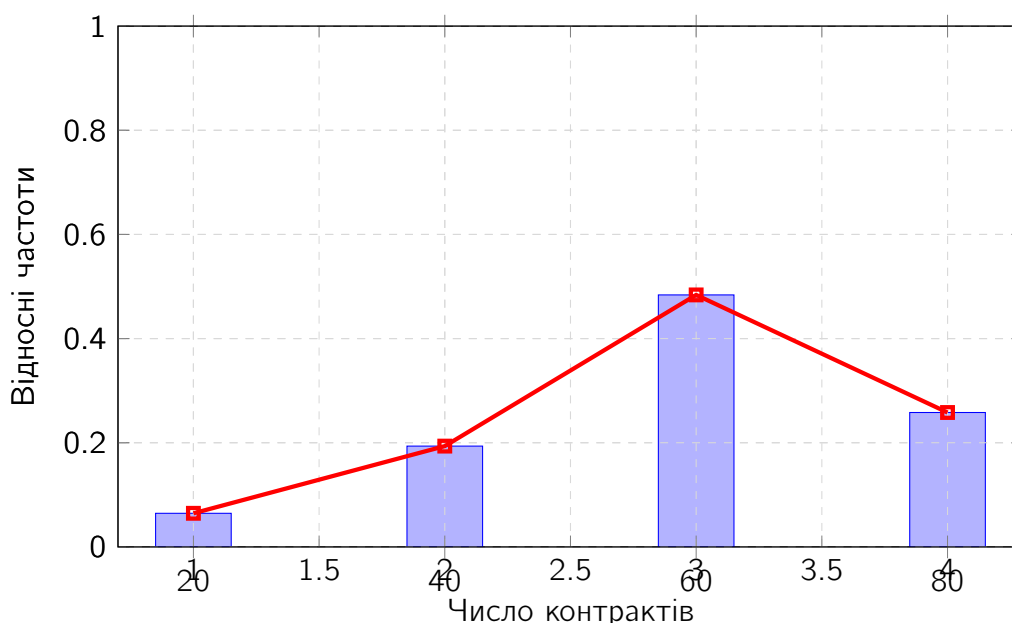
Розв'язання. Побудуємо полігон частот. Для цього спочатку знайдемо відносні частоти. Загальна кількість укладених контрактів:

$$N = 20 \cdot 10 + 40 \cdot 15 + 60 \cdot 25 + 80 \cdot 10 = 200 + 600 + 1500 + 800 = 3100.$$

Таким чином, відносні частоти дорівнюють:

$$\begin{aligned} p_1 &= \frac{10 \cdot 20}{3100} = \frac{2}{31}, \\ p_2 &= \frac{15 \cdot 40}{3100} = \frac{6}{31}, \\ p_3 &= \frac{25 \cdot 60}{3100} = \frac{15}{31}, \\ p_4 &= \frac{10 \cdot 80}{3100} = \frac{8}{31}. \end{aligned}$$

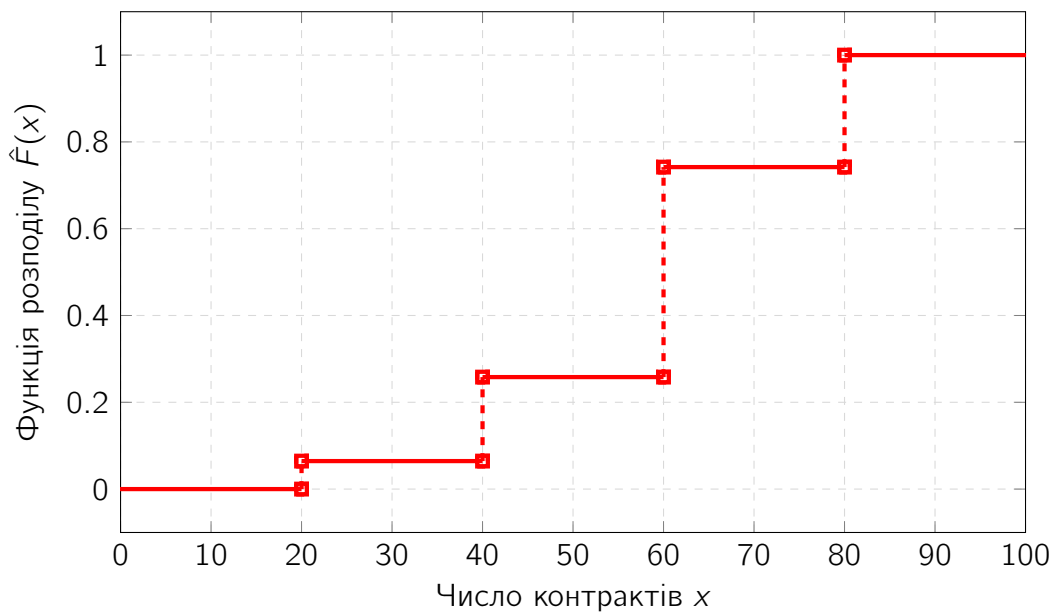
Тепер можемо побудувати полігон частот:



Тепер побудуємо вибірку функцію розподілу. Для цього відсортуємо дані за зростанням, а потім знайдемо відносні частоти для кожного інтервалу:

$$\hat{F}(x) = \begin{cases} 0, & x \leq 20, \\ \frac{2}{31}, & 20 < x \leq 40, \\ \frac{2}{31} + \frac{6}{31}, & 40 < x \leq 60, \\ \frac{2}{31} + \frac{6}{31} + \frac{15}{31}, & 60 < x \leq 80, \\ 1, & x > 80. \end{cases}$$

Тепер можемо побудувати вибірку функцію розподілу:



Гістограму сил намалювати не вистачило :(

Знайдемо вибіркове середнє:

$$\hat{\mu} = \frac{10 \cdot 20 + 15 \cdot 40 + 25 \cdot 60 + 10 \cdot 80}{10 + 15 + 25 + 10} = \frac{3100}{60} = \frac{310}{6} \approx 51.7.$$

Тепер знайдемо вибірку дисперсію:

$$\begin{aligned} \bar{\sigma}^2 &= \frac{1}{60} (10 \cdot 20^2 + 15 \cdot 40^2 + 25 \cdot 60^2 + 10 \cdot 80^2) - \hat{\mu}^2 \\ &\approx 3033.33 - 2669.44 \approx 363.89. \end{aligned}$$

Незміщена оцінка дисперсії:

$$\hat{\sigma}^2 = \frac{n}{n-1} \cdot \bar{\sigma}^2 \approx \frac{60}{59} \cdot 363.89 \approx 370.$$

2.2 Вправа 2. Час на виготовлення деталі

Умова Задачі 2.2. Нижче наведені дані про час, витрачений робочими на виготовлення однієї деталі. Побудувати вибірку функцію розподілу, гістограму вибірки та полігон частот. Знайти вибіркове середнє, вибірккову дисперсію і незміщену оцінку дисперсії.

Інтервали часу, хвил.	[4, 4.4)	[4.4, 4.8)	[4.8, 5.2)	[5.2, 5.6)	[5.6, 6.0)
Кількість робочих	5	8	21	31	19

Розв'язання. Побудуємо вибірку функцію розподілу. Знайдемо відносні частоти:

$$\begin{aligned} p_1 &= \frac{5}{84}, \\ p_2 &= \frac{8}{84}, \\ p_3 &= \frac{21}{84}, \\ p_4 &= \frac{31}{84}, \\ p_5 &= \frac{19}{84}. \end{aligned}$$

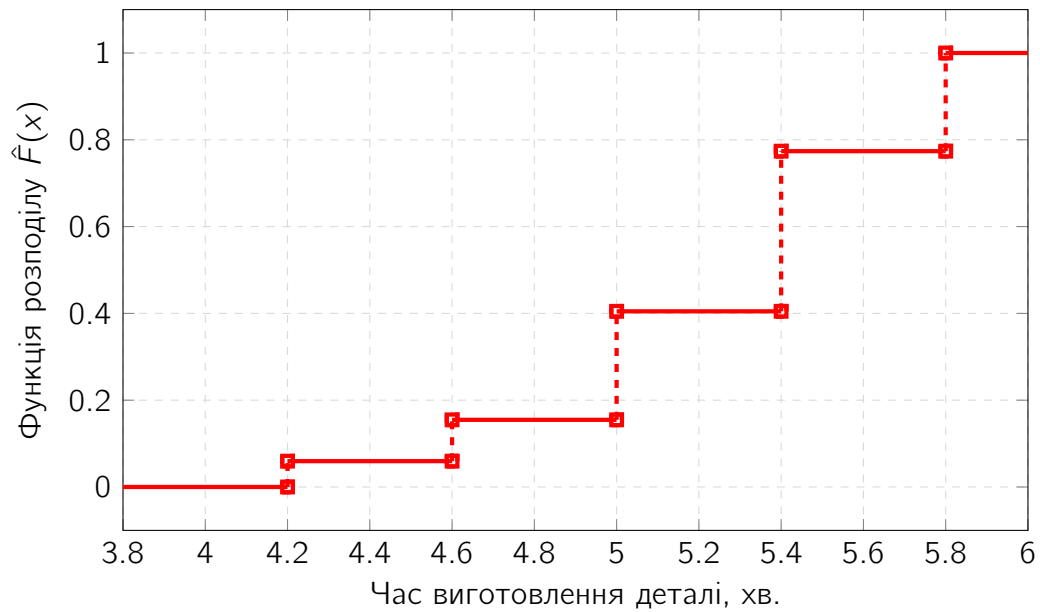
Відповідні середини інтервалів:

$$\begin{aligned} z_1 &= \frac{4 + 4.4}{2} = 4.2, \\ z_2 &= \frac{4.4 + 4.8}{2} = 4.6, \\ z_3 &= \frac{4.8 + 5.2}{2} = 5, \\ z_4 &= \frac{5.2 + 5.6}{2} = 5.4, \\ z_5 &= \frac{5.6 + 6}{2} = 5.8. \end{aligned}$$

Тепер можемо побудувати вибірку функцію розподілу. Для цього відсортуємо дані за зростанням, а потім знайдемо відносні частоти для кожного інтервалу:

$$\hat{F}(x) = \begin{cases} 0, & x \leq 4.2, \\ \frac{5}{84}, & 4.2 < x \leq 4.6, \\ \frac{5}{84} + \frac{8}{84}, & 4.6 < x \leq 5.0, \\ \frac{5}{84} + \frac{8}{84} + \frac{21}{84}, & 5.0 < x \leq 5.4, \\ \frac{5}{84} + \frac{8}{84} + \frac{21}{84} + \frac{31}{84}, & 5.4 < x \leq 5.8, \\ 1, & x > 5.8. \end{cases}$$

Тепер можемо побудувати вибірку функцію розподілу:



Знайдемо вибіркове середнє:

$$\hat{\mu} = \frac{5 \cdot 4.2 + 8 \cdot 4.6 + 21 \cdot 5 + 31 \cdot 5.4 + 19 \cdot 5.8}{5 + 8 + 21 + 31 + 19} = \frac{440.4}{84} \approx 5.25.$$

Тепер знайдемо вибірккову дисперсію:

$$\begin{aligned} \bar{\sigma}^2 &= \frac{1}{84} (5 \cdot 4.2^2 + 8 \cdot 4.6^2 + 21 \cdot 5^2 + 31 \cdot 5.4^2 + 19 \cdot 5.8^2) - \hat{\mu}^2 \\ &\approx 27.69 - 27.56 \approx 0.13. \end{aligned}$$

Незміщена оцінка дисперсії:

$$\hat{\sigma}^2 = \frac{84}{83} \cdot 0.13 \approx 0.132.$$