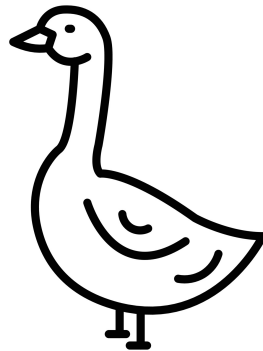


# Розрахункова робота № 1

з Математичної Статистики

студента КА-02  
Романовича Володимира  
Варіант 9



Київ 2022

## Завдання розрахункової роботи

1. Побудувати варіаційний (дискретний або інтервальний) ряд наданої вибірки.
2. Зробити графічне зображення вибірки.
3. Побудувати емпіричну функцію розподілу.
4. Знайти незміщену оцінку математичного сподівання та дисперсії.
5. Обчислити значення вибіркової медіани, моди, асиметрії.
6. Висунути гіпотезу про розподіл, за яким отримано вибірку.
7. Знайти точкові оцінки параметрів гіпотетичного закону розподілу та перевірити їх властивості.
8. Перевірити за допомогою критерію значущості (Пірсона) гіпотезу про розподіл з рівнем  $\alpha = 0,05$ .
9. Знайти довірчий інтервал для параметрів гіпотетичного закону розподілу, взявши рівень надійності 0.95.
10. Висновки.

### Задана конкретна реалізація вибірки:

$\vec{x} = (5.04, 8.16, 4.32, 4.29, 4.26, 4.33, 6.17, 4.63, 4.28, 4.37, 7.74, 5.19, 4.19, 4.71, 5.29, 5.53, 5.32, 4.18, 4.47, 4.94, 5.51, 4.66, 7.62, 4.23, 7.3, 5.95, 4.51, 5.46, 4.81, 4.27, 7.58, 5.9, 6.02, 5.72, 6.33, 5.12, 4.6, 4.51, 6.56, 5.55, 5.48, 4.27, 4.99, 4.3, 5.84, 5.72, 5.02, 4.7, 7.4, 4.18, 5.35, 4.98, 5.58, 5.1, 5.18, 5.67, 4.95, 6.09, 5.01, 4.62, 4.47, 7.54, 7.0, 4.83, 6.43, 4.52, 4.66, 5.69, 5.5, 5.24, 5.86, 4.27, 5.58, 4.73, 4.39, 4.19, 13.09, 4.22, 7.29, 9.24, 5.95, 4.18, 6.15, 4.37, 6.16, 4.68, 6.46, 5.05, 4.65, 7.76, 5.12, 5.64, 5.52, 6.38, 5.44, 4.91, 7.84, 7.22, 4.2, 5.39)$

Можемо побачити що у заданій нам реалізації вибірки майже всі значення унікальні, тому припустимо, що генеральна сукупність, за якою задано конкретну реалізацію вибірки, є неперервною випадковою величиною. Тому доцільно буде використовувати інтервальний варіаційний ряд.

### 1. Побудувати інтервальний варіаційний ряд наданої вибірки

Скористаємося правилом Стерджеса щоб знати на скільки інтервалів ділити відрізок  $[x_{min}, x_{max}]$ :  
 $k = 1 + [3.322 \lg 100] = 7$ , розмах вибірки  $R = x_{max} - x_{min} = 13.09 - 4.28 = 8.81$ , тому поділимо відрізок  $[4.18, 13.09]$  на 7 рівних напівінтервалів довжиною  $\frac{R}{k} = \frac{8.81}{7} \approx 1.2585$

Табл. 1: Інтервальний варіаційний ряд конкретної реалізації вибірки з інтервалами рівної довжини

Інтервал $\Delta_i$	[4.18; 5.45)	[5.45; 6.73)	[6.73; 8)	[8; 9.27)	[9.27; 10.54)	[10.54; 11.82)	[11.82; 13.09]
Частоти $n_i$	57	29	11	2	0	0	1
Частоти $\omega_i$	0.57	0.29	0.11	0.02	0	0	0.01

Можемо замітити що тільки 3 координати нашої реалізації вибірки потрапили в відрізок  $[8; 13.09]$ , при чому в півінтервал  $[9.27; 11.82)$  не потрапило жодної координати. Тому я вважаю доцільним об'єднати останні 4 інтервали в один довжиною 5.09 та поділити півінтервал  $[4.18; 8)$  вже на 6 рівних інтервалів довжиною приблизно 0.64. Побудуємо такий варіаційний ряд. Надалі використовуватимемо саме його

Табл. 2: Інтервальний варіаційний ряд конкретної реалізації вибірки

Інтервал $\Delta_i$	[4.18; 4.82)	[4.82; 5.45)	[5.45; 6.09)	[6.09; 6.73)	[6.73; 7.36)	[7.36; 8)	[8; 13.09]
Частоти $n_i$	36	21	20	9	4	7	3
Частоти $\omega_i$	0.36	0.21	0.20	0.09	0.04	0.07	0.03

## 2. Зробити графічне зображення вибірки

Графічним зображенням конкретної реалізації вибірки при неперервно розподіленій ГС є гістограма, що складається з прямокутників, побудованих на інтервалах  $\Delta_i$  з висотами  $h_i = \frac{\omega_i}{l(\Delta_i)}$

Додамо до таблиці значення цих висот  $h_i$ :

Інтервал $\Delta_i$	[4.18; 4.82)	[4.82; 5.45)	[5.45; 6.09)	[6.09; 6.73)	[6.73; 7.36)	[7.36; 8)	[8; 13.09]
Частоти $n_i$	36	21	20	9	4	7	3
Частоти $\omega_i$	0.36	0.21	0.20	0.09	0.04	0.07	0.03
Висоти $h_i$	0.565	0.33	0.314	0.141	0.063	0.11	0.006

Побудуємо тепер гістограму, яка і буде геометричною інтерпретацією даної вибірки:

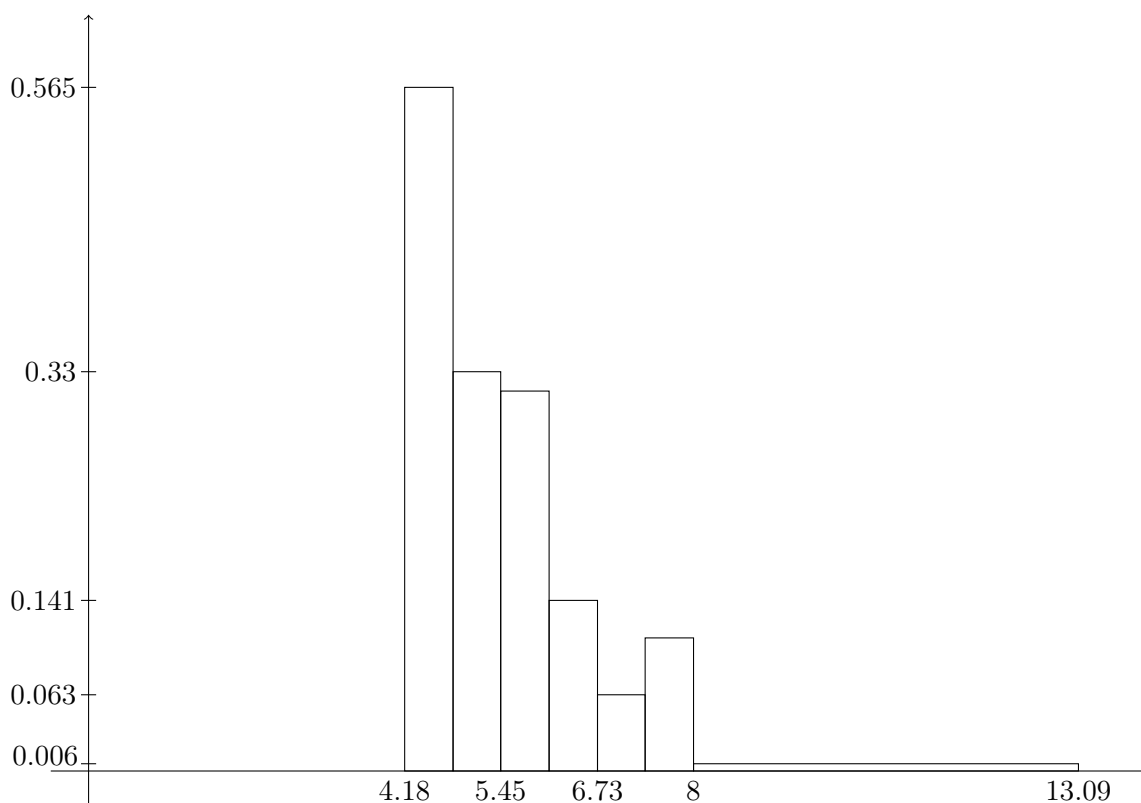


Рис. 1: Гістограма конкретної реалізації вибірки

Також побудуємо емпіричну функцію розподілу нашої конкретної реалізації, та її графік

$$\begin{aligned}
 (F_n^*)_{\text{зн}}(x) &= \begin{cases} 0, & x \leq x_{\min} \\ \omega_1^{\text{нак}}, & x \in (x_{\min}; t_1] \\ \omega_2^{\text{нак}}, & x \in (t_1; t_2] \\ \omega_3^{\text{нак}}, & x \in (t_2; t_3] \\ \omega_4^{\text{нак}}, & x \in (t_3; t_4] \\ \omega_5^{\text{нак}}, & x \in (t_4; t_5] \\ \omega_6^{\text{нак}}, & x \in (t_5; t_6] \\ 1, & x > t_6 \end{cases} = \begin{cases} 0, & x \leq 4.18 \\ 0.36, & x \in (4.18; 4.82] \\ 0.36 + 0.21, & x \in (4.82; 5.45] \\ 0.36 + 0.21 + 0.2, & x \in (5.45; 6.09] \\ 0.36 + 0.21 + 0.2 + 0.09, & x \in (6.09; 6.73] \\ 0.36 + 0.21 + 0.2 + 0.09 + 0.04, & x \in (6.73; 7.36] \\ 0.36 + 0.21 + 0.2 + 0.09 + 0.04 + 0.07, & x \in (7.36; 8] \\ 1, & x > 8 \end{cases} = \\
 &= \begin{cases} 0, & x \leq 4.18 \\ 0.36, & x \in (4.18; 4.82] \\ 0.57, & x \in (4.82; 5.45] \\ 0.77, & x \in (5.45; 6.09] \\ 0.86, & x \in (6.09; 6.73] \\ 0.9, & x \in (6.73; 7.36] \\ 0.97, & x \in (7.36; 8] \\ 1, & x > 8 \end{cases}
 \end{aligned}$$

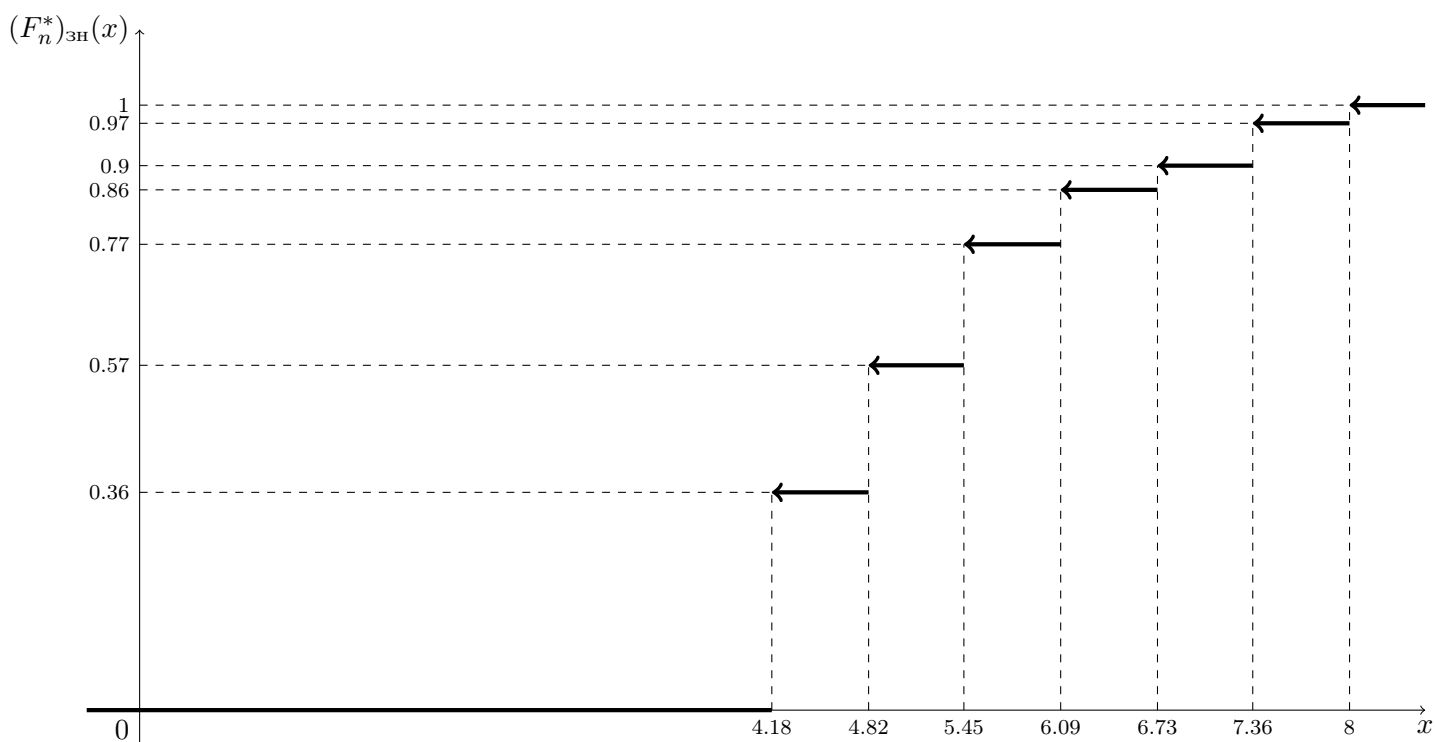


Рис. 2: Емпірична функція розподілу конкретної реалізації вибірки

Оскільки ми зробили припущення що наша ГС – неперервна, функція розподілу такої ГС також неперервна. Тому побудуємо також кумулятивну криву

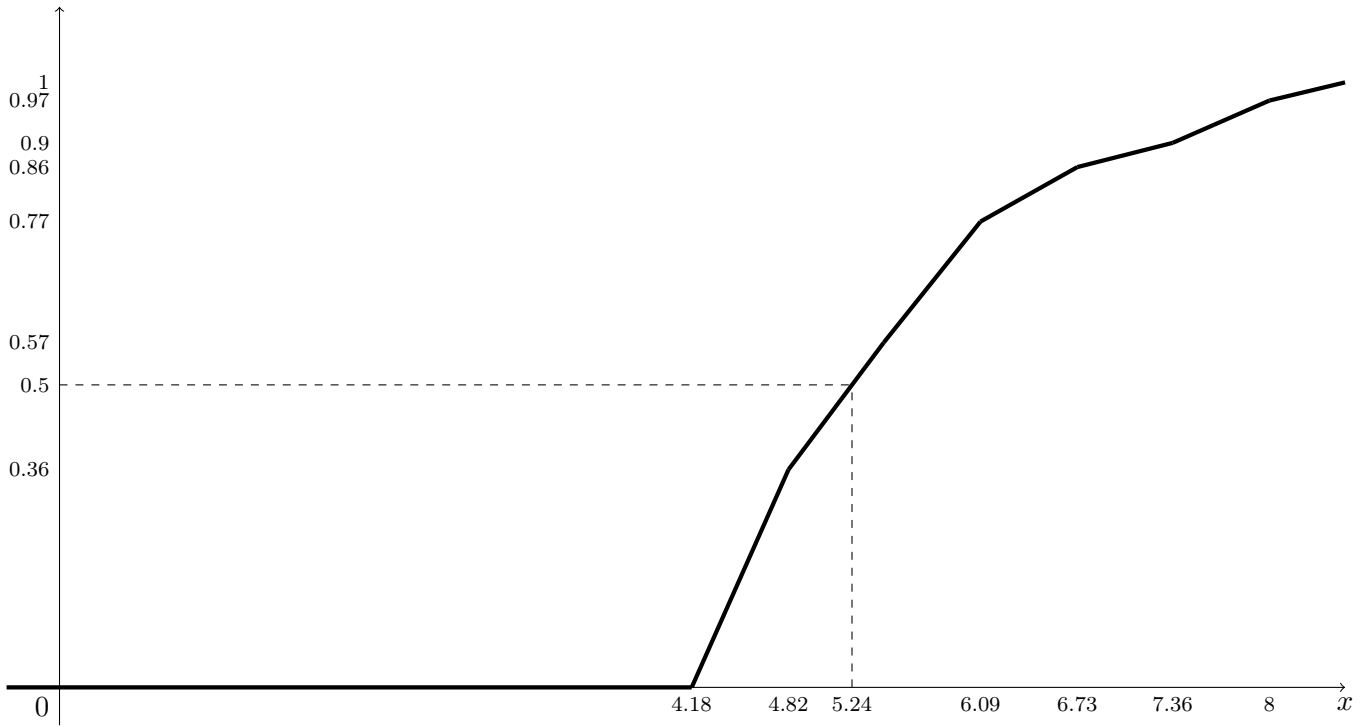


Рис. 3: Кумулятивна крива конкретної реалізації вибірки

#### 4. Знайти незміщену оцінку математичного сподівання та дисперсії

Нагадаємо що оцінка  $\theta^*$  параметру  $\theta$  називається незміщеною, якщо  $\mathbb{E}\theta_n^* = \theta$

Як оцінку математичного сподівання виберемо вибіркове середнє:  $\bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k$

Перевіримо незміщеність цієї оцінки:  $\mathbb{E}\bar{\xi} \stackrel{?}{=} \mathbb{E}\xi$

$$\mathbb{E}\bar{\xi} = \mathbb{E}\left(\frac{1}{n} \sum_{k=1}^n \xi_k\right) = \frac{1}{n} \mathbb{E}\left(\sum_{k=1}^n \xi_k\right) \stackrel{\text{(власивість)}}{=} \frac{1}{n} \sum_{k=1}^n \mathbb{E}\xi_k \stackrel{\text{(всі } \xi_k \text{ розподілені як } \xi)}{=} \frac{1}{n} \cdot n \cdot \mathbb{E}\xi = \mathbb{E}\xi$$

Отже за означенням  $\bar{\xi}$  – незміщена оцінка  $\mathbb{E}\xi$

Як оцінку дисперсії виберемо виправлену вибірккову дисперсію:  $D^{**}\xi = \frac{1}{n-1} \sum_{k=1}^n (\xi_k - \bar{\xi})^2$

Перевіримо незміщеність цієї оцінки:  $\mathbb{E}(D^{**}\xi) \stackrel{?}{=} D\xi$

$$\begin{aligned} \mathbb{E}(D^{**}\xi) &= \mathbb{E}\left(\frac{1}{n-1} \sum_{k=1}^n (\xi_k - \bar{\xi})^2\right) = \frac{1}{n-1} \cdot \mathbb{E}\left(\sum_{k=1}^n ((\xi_k - \mathbb{E}\xi) - (\bar{\xi} - \mathbb{E}\xi))^2\right) = \\ &= \frac{1}{n-1} \cdot \mathbb{E}\left(\sum_{k=1}^n (\xi_k - \mathbb{E}\xi)^2 - 2(\bar{\xi} - \mathbb{E}\xi) \cdot \underbrace{\sum_{k=1}^n (\xi_k - \mathbb{E}\xi)}_{n(\bar{\xi} - \mathbb{E}\xi)} + n(\bar{\xi} - \mathbb{E}\xi)^2\right) = \\ &= \frac{1}{n-1} \cdot \mathbb{E}\left(\sum_{k=1}^n (\xi_k - \mathbb{E}\xi)^2 - 2n(\bar{\xi} - \mathbb{E}\xi)^2 + n(\bar{\xi} - \mathbb{E}\xi)^2\right) = \frac{1}{n-1} \cdot \mathbb{E}\left(\sum_{k=1}^n (\xi_k - \mathbb{E}\xi)^2 - n(\bar{\xi} - \mathbb{E}\xi)^2\right) = \\ &\stackrel{(1)}{=} \frac{1}{n-1} \cdot \mathbb{E}\left(\sum_{k=1}^n (\xi_k - \mathbb{E}\xi)^2 - n(\bar{\xi} - \mathbb{E}\bar{\xi})^2\right) \stackrel{(2)}{=} \frac{1}{n-1} \cdot (nD\xi - nD\bar{\xi}) \stackrel{(3)}{=} \frac{1}{n-1} \cdot (nD\xi - D\xi) = D\xi \end{aligned}$$

Пояснення переходу (1): Вище доведено незміщеність  $\bar{\xi}$  як оцінки  $\mathbb{E}\xi$ , тому  $\mathbb{E}\xi = \mathbb{E}\bar{\xi}$

Пояснення переходу (2): Всі  $\xi_k$  розподілені як  $\xi$ , тому  $\mathbb{E}(\xi_k - \mathbb{E}\xi)^2 = \mathbb{E}(\xi - \mathbb{E}\xi)^2 = D\xi$ , отже

$$\sum_{k=1}^n \mathbb{E}(\xi_k - \mathbb{E}\xi)^2 = n \cdot D\xi; \quad \bar{\xi} - \text{випадкова величина, тому } \mathbb{E}(\bar{\xi} - \mathbb{E}\bar{\xi})^2 = D\bar{\xi}$$

Пояснення переходу (3):  $D\bar{\xi} = D\left(\frac{1}{n} \sum_{k=1}^n \xi_k\right) = \frac{1}{n^2} D\left(\sum_{k=1}^n \xi_k\right) = \frac{1}{n} D\xi$ , останній перехід можливий за рахунок того що координати випадкової вибірки однаково розподілені і незалежні  
Отримали що за означенням  $D^{**}\xi$  – незміщена оцінка дисперсії

## 5. Обчислити значення вибіркової медіани, моди, асиметрії

Для початку обчислимо значення вибіркового середнього та виправленої вибіркової дисперсії:

$$\text{Вибіркове середнє } \bar{\xi} = \frac{1}{n} \sum_{k=1}^n \xi_k$$

$$\text{Значення вибіркового середнього: } \bar{x} = \frac{1}{n} \sum_{k=1}^n x_k = \frac{1}{100} \sum_{k=1}^{100} x_k = 5.4986$$

$$\text{Виправлена вибіркова дисперсія } D^{**}\xi = \frac{1}{n-1} \sum_{k=1}^n (\xi_k - \bar{\xi})^2$$

$$\text{Значення виправленої вибіркової дисперсії: } (D^{**}\xi)_{\text{зн}} = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2 = \frac{1}{99} \sum_{k=1}^{100} (x_k - 5.4986)^2 \approx 1.784$$

**Вибіркова мода**

$$(Mo_1^*\xi)_{\text{зн}} = y_i + l(\Delta_i) \cdot \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} = 4.18 + 0.64 \cdot \frac{36}{36 + (36 - 21)} \approx 4.632$$

Також знайдемо значення вибіркової моди графічно, для цього візьмемо графік найвищого стовпця гістограми, та його сусідніх стовпців (оскільки найвищий стовпчик у нас перший, то сусідній буде тільки 1):



Цим способом отримуємо  $(Mo_2^*\xi)_{\text{зн}} \approx 4.63$

**Вибіркова медіана**

$$(Me_1^*\xi)_{\text{зн}} = y_i + \frac{l(\Delta_i)}{n_i} \cdot \left(\frac{n}{2} - \sum_{k=1}^{i-1} n_k\right) = 4.82 + \frac{0.64}{21} \cdot \left(\frac{100}{2} - 36\right) \approx 5.247$$

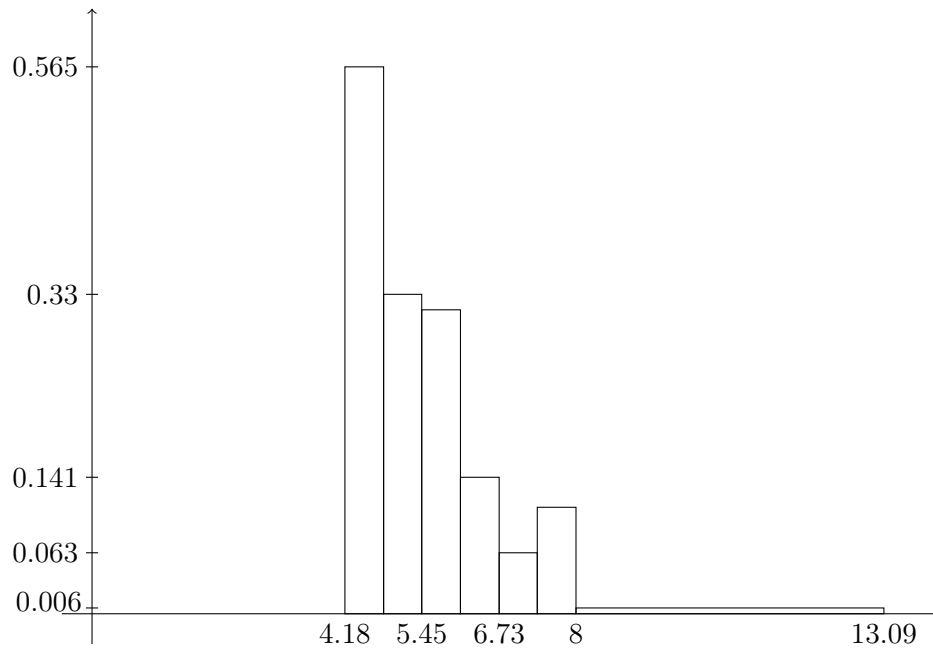
Також можна знайти значення вибіркової медіани за кумулятивною кривою, абсциса точки в якій кумулянта набуває значення 0.5 і буде значенням вибіркової медіани, в нашому випадку  $(Me_2^*\xi)_{\text{зн}} \approx 5.24$

$$\text{Вибіркова асиметрія } As^*\xi = \frac{\frac{1}{n} \sum_{k=1}^n (\xi_k - \bar{\xi})^3}{(D^*\xi)^{3/2}}$$

$$(As^*\xi)_{\text{зн}} = \frac{\frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^3}{(D^{**}\xi)_{\text{зн}}^{3/2}} \approx \frac{\frac{1}{100} \sum_{k=1}^{100} (x_k - 5.4986)^3}{1.784^{3/2}} \approx 2.3215$$

## 6. Висунути гіпотезу про розподіл, за яким отримано вибірку

Для висування гіпотези щодо розподілу, подивимося ще раз на гістограму, співставлену нашій конкретній реалізації вибірки.



Оскільки гістограма є своєрідним аналогом функції щільності ГС, з вигляду гістограми можемо припустити що на півосі  $[4.18; +\infty)$  щільність є спадною, при чому спадає достатньо різко і при  $x > 8$  ця щільність є дуже малою (лиш 3 координати  $\vec{x}$  потрапили в цю піввісь). Це характерно для щільності показникового розподілу, при чому оскільки найменше значення  $\vec{x}$  це 4.18, то якщо ГС має експоненційний розподіл, майже напевно він із зсувом. (оскільки на лекційних заняттях не було функції розподілу та числових характеристик показникового розподілу із зсувом, їх виведено у [додатку](#)). Оскільки значення параметру зсуву є найменшим значенням, яке може приймати ВВ, при чому це і найімовірніше значення (мода ВВ), можна зробити припущення що значення параметру зсуву близьке до мінімального значення конкретної реалізації вибірки, але не більше за нього (щільність при  $x < \theta$  нульова). Тому припустимо, що значення параметру зсуву  $\theta \approx \min_{1 \leq k \leq n} x_k = 4.18$

Подивимось на числові характеристики такого розподілу:

Нехай  $\eta \sim \text{Exp}(\lambda, \theta)$ ,  $D\eta = \lambda^2$ . Ми маємо  $(D^{**}\xi)_{\text{зн}} \approx 1.784$ ,  $\sqrt{(D^{**}\xi)_{\text{зн}}} \approx 1.336$

$\sqrt{D\eta} = \lambda$ , отже можемо припустити що значення  $\lambda \approx 1.336$

Перевіримо тепер інші числові характеристики припустивши що  $\eta \sim \text{Exp}(1.336, 4.18)$

$E\eta = \lambda + \theta \approx 1.336 + 4.18 = 5.516$ , помітимо що ми отримали значення, близьке до  $\bar{x} = 5.4986$

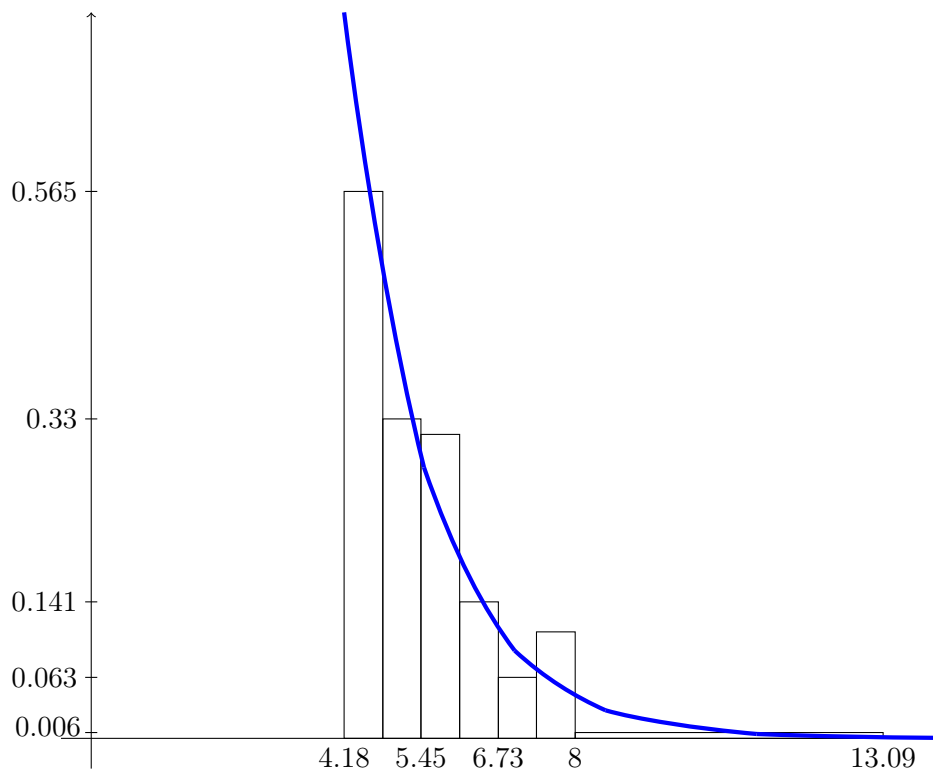
Також ми знаємо що у показниковому розподілі із зсувом, мода є значенням параметру зсуву.

$(Mo_1^*\xi)_{\text{зн}} = 4.632$ , що є близьким до припущеного значення  $\theta \approx 4.18$

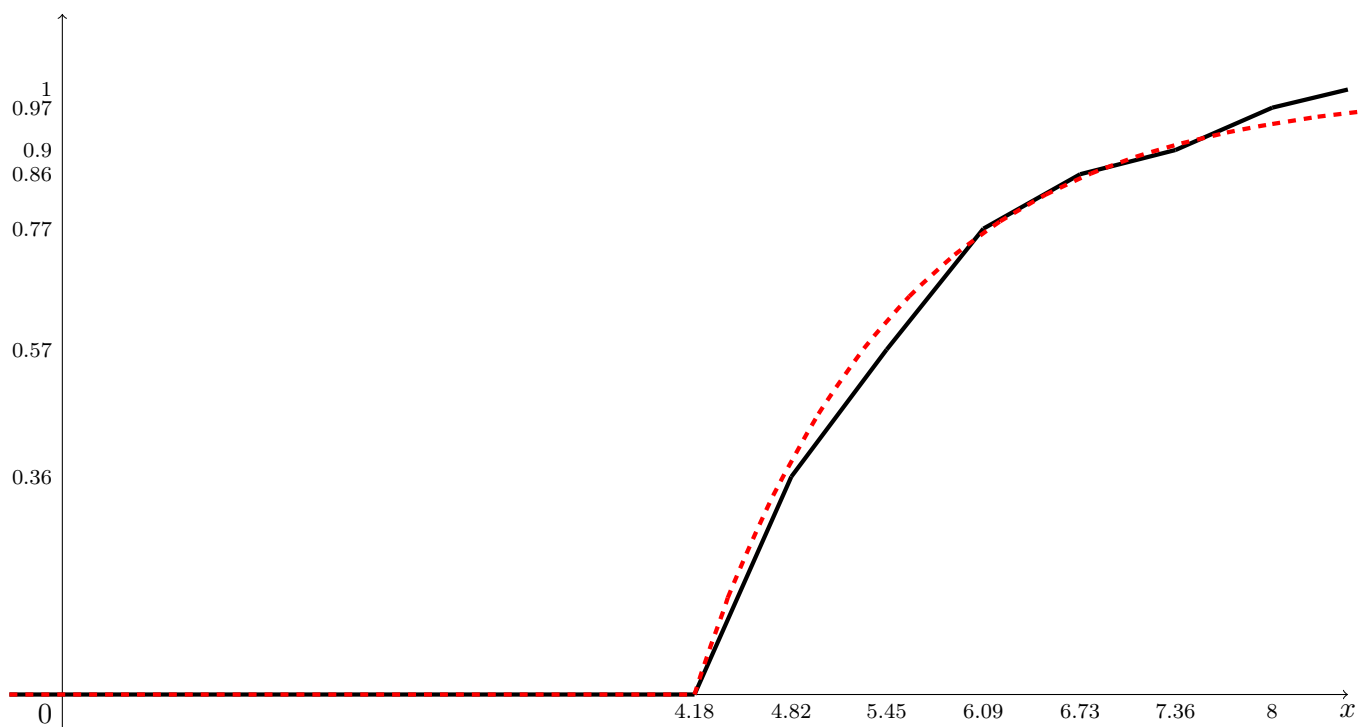
$Me\eta = \lambda \ln 2 + \theta \approx 5.106$ . Нагадаємо що  $(Me_1^*\xi)_{\text{зн}} \approx 5.247$ .

Асиметрія гіпотетичного розподілу  $As\eta = 2$ , значення вибіркової асиметрії – 2.3215

Побудуємо тепер графік щільності  $\eta \sim \text{Exp}(1.336, 4.18)$  поверх нашої гістограми.



Як бачимо з малюнку, крива щільності достатньо близька до стовпчиків гістограми. Порівняймо також нашу кумулятивну криву із графіком функції розподілу  $\eta$



Знову ж таки отримали достатньо близькі графіки. Враховуючи все вище сказане вважаю доцільним ввести гіпотезу  $H_0 : \xi \sim Exp(\lambda, \theta)$ , знайти точкові оцінки та перевірити цю гіпотезу щодо розподілу генеральної сукупності критерієм Пірсона.

## 7. Знайти точкові оцінки параметрів гіпотетичного закону розподілу та перевірити їх властивості



Для знаходження точкових оцінок використаємо метод моментів та метод максимальної правдоподібності.

**Метод моментів:**

$$\begin{cases} \mathbb{E}\xi = \lambda + \theta \\ D\xi = \lambda^2 \end{cases} \implies \begin{cases} \theta = \mathbb{E}\xi - \lambda \\ \lambda = \sqrt{D\xi} \end{cases} \implies \begin{cases} \theta = \mathbb{E}\xi - \sqrt{D\xi} \\ \lambda = \sqrt{D\xi} \end{cases} \implies \begin{cases} \theta_{MM}^* = \bar{\xi} - \sqrt{D^{**}\xi} \\ \lambda_{MM}^* = \sqrt{D^{**}\xi} \end{cases}$$

**Метод максимальної правдоподібності:**

Нагадаємо, що  $\ln \mathcal{L}_{Exp}(\vec{x}, \lambda, \theta) = -n \ln \lambda - \frac{1}{\lambda} \sum_{k=1}^n x_k + \frac{n\theta}{\lambda}$

Згідно методу нам потрібно знайти  $\arg\max_{\lambda, \theta} \ln \mathcal{L}_{Exp}(\vec{x}, \lambda, \theta)$ :

Необхідна умова екстремуму:

$$\begin{cases} \frac{\partial \ln \mathcal{L}_{Exp}(\vec{x}, \lambda, \theta)}{\partial \lambda} = -\frac{n}{\lambda} + \frac{1}{\lambda^2} \sum_{k=1}^n x_k - \frac{n\theta}{\lambda^2} = 0 \\ \frac{\partial \ln \mathcal{L}_{Exp}(\vec{x}, \lambda, \theta)}{\partial \theta} = \frac{n}{\lambda} \neq 0 \end{cases}$$

Отже екстремуму функції двох змінних немає.

Але ми бачимо що  $\frac{\partial \ln \mathcal{L}_{Exp}(\vec{x}, \lambda, \theta)}{\partial \theta} > 0$ , отже  $\ln \mathcal{L}_{Exp}(\vec{x}, \lambda, \theta)$  зростає із зростанням  $\theta$ .

Проте  $\theta$  обмежена зверху значенням  $\min_{1 \leq k \leq n} x_k$ , звідси  $\ln \mathcal{L}_{Exp}(\vec{x}, \lambda, \theta)$  набуватиме найбільшого значення при

$$\theta = \min_{1 \leq k \leq n} x_k.$$

Залишається знайти при якому  $\lambda$   $\ln \mathcal{L}_{Exp}(\vec{x}, \lambda, \theta)$  набуватиме найбільшого значення.

$$\frac{\partial \ln \mathcal{L}_{Exp}(\vec{x}, \lambda, \theta)}{\partial \lambda} = -\frac{n}{\lambda} + \frac{n}{\lambda^2} \bar{x} - \frac{n\theta}{\lambda^2} = \frac{n}{\lambda^2} (-\lambda + \bar{x} - \theta) = 0 \implies \lambda_{кр} = \bar{x} - \theta$$

$$\left. \frac{\partial^2 \ln \mathcal{L}_{Exp}(\vec{x}, \lambda, \theta)}{\partial \lambda^2} \right|_{\lambda=\bar{x}-\theta} = \left( \frac{n}{\lambda^2} - \frac{2n(\bar{x} - \theta)}{\lambda^3} \right) \Big|_{\lambda=\bar{x}-\theta} = -\frac{n}{(\bar{x} - \theta)^2} < 0$$

Отже  $\lambda_{max} = \bar{x} - \theta$ ,  $\theta_{max} = \min_{1 \leq k \leq n} x_k$

Звідси  $\arg\max_{\lambda, \theta} \ln \mathcal{L}_{Exp}(\vec{x}, \lambda, \theta) = \ln \mathcal{L}_{Exp}(\vec{x}, \bar{x} - \theta, \min x_k)$ , отримуємо оцінки:

$$\begin{cases} \lambda_{ММП}^{**} = \bar{\xi} - \theta \\ \theta_{ММП}^{**} = \min_{1 \leq k \leq n} \xi_k \end{cases}$$

Перевіримо властивості оцінок, отриманих методом максимальної правдоподібності:

① Перевірка властивостей  $\lambda^{**} = \bar{\xi} - \theta$ :

**Незміщеність:**

$$\mathbb{E}\lambda^{**} = \mathbb{E} \left( \frac{1}{n} \sum_{k=1}^n \xi_k - \theta \right) = \frac{1}{n} \mathbb{E} \left( \sum_{k=1}^n \xi_k \right) - \mathbb{E}\theta = \mathbb{E}\xi - \theta = \lambda + \theta - \theta = \lambda$$

Отже за означенням оцінка є незміщеною

**Конзистентність:**

Використаємо достатню умову конзистентності для незміщених оцінок:

$$D\lambda^{**} = D(\bar{\xi} - \theta) = D\bar{\xi} = D \left( \frac{1}{n} \sum_{k=1}^n \xi_k \right) = \frac{1}{n^2} \sum_{k=1}^n D\xi = \frac{1}{n^2} \cdot n D\xi = \frac{1}{n} \lambda^2 \rightarrow 0, n \rightarrow \infty$$

Отже за достатньою умовою оцінка є конзистентною

**Ефективність:**

Найзручніше доводити ефективність за наслідком з нерівності Рао-Крамера, використаємо його:

$$\frac{\partial \ln \mathcal{L}_{Exp}(\vec{\xi}, \lambda, \theta)}{\partial \lambda} = -\frac{n}{\lambda} + \frac{n}{\lambda^2} \bar{\xi} - \frac{n\theta}{\lambda^2} = \frac{n}{\lambda^2} (-\lambda + \bar{\xi} - \theta) = \frac{n}{\lambda^2} (\lambda^{**} - \lambda) = C(n, \lambda) \cdot (\lambda^{**} - \lambda)$$

Отже за наслідком з нерівності Рао-Крамера, оцінка є ефективною.

#### Асимптотична нормальність:

Центруємо і нормуємо нашу випадкову величину  $\lambda^{**}$ :

$$\frac{\lambda^{**} - \mathbb{E}\lambda^{**}}{\sqrt{D\lambda^{**}}} = \frac{\lambda^{**} - \lambda}{\sqrt{\frac{\lambda^2}{n}}} = \frac{\sqrt{n} \left( \frac{1}{n} \sum_{k=1}^n \xi_k - \theta - (\mathbb{E}\xi - \theta) \right)}{\lambda} = \frac{\frac{1}{n} \sqrt{n} \left( \sum_{k=1}^n (\xi_k - \mathbb{E}\xi) \right)}{\sqrt{D\xi}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \left( \frac{\xi_k - \mathbb{E}\xi}{\sqrt{D\xi}} \right)$$

За наслідком з теореми Ляпунова:

$$\frac{\lambda^{**} - \mathbb{E}\lambda^{**}}{\sqrt{D\lambda^{**}}} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \left( \frac{\xi_k - \mathbb{E}\xi}{\sqrt{D\xi}} \right) \xrightarrow{F} \eta \sim N(0, 1), n \rightarrow \infty$$

Отже за означенням оцінка є асимптотично нормальною.

#### ② Перевірка властивостей $\theta^{**} = \min_{1 \leq k \leq n} \xi_k$ :

Для початку знайдемо закон розподілу мінімуму  $\xi_k$ :

$$f_{\theta^{**}}(x) = n(1 - F_{\xi}(x))^{n-1} f_{\xi}(x) = n \left( 1 - 1 + e^{-\frac{1}{\lambda}(x-\theta)} \right)^{n-1} \cdot \frac{1}{\lambda} e^{-\frac{1}{\lambda}(x-\theta)} = \frac{n}{\lambda} e^{-\frac{n}{\lambda}(x-\theta)}, x \geq \theta$$

З щільності розподілу мінімуму бачимо, що  $\theta^{**} \sim \text{Exp}\left(\frac{\lambda}{n}, \theta\right)$

#### Незміщеність:

$$\mathbb{E}\theta^{**} = \frac{\lambda}{n} + \theta$$

Бачимо що така оцінка є тільки асимптотично незміщеною, розглянемо оцінку  $\theta_1^* = \min \xi_k - \frac{\lambda}{n}$ :

$$\mathbb{E}\theta_1^* = \mathbb{E} \left( \theta^{**} - \frac{\lambda}{n} \right) = \mathbb{E}\theta^{**} - \frac{\lambda}{n} = \theta + \frac{\lambda}{n} - \frac{\lambda}{n} = \theta$$

Така оцінка є незміщеною, тому надалі розглядатимемо саме її.

#### Конзистентність:

Знайдемо  $D\theta_1^*$

$$D\theta_1^* = D \left( \theta^{**} - \frac{\lambda}{n} \right) = D\theta^{**} = \left( \frac{\lambda}{n} \right)^2 \rightarrow 0, n \rightarrow \infty$$

Отже за достатньою умовою оцінка є конзистентною

#### Ефективність:

Використаємо нерівність Рао-Крамера:

$$\mathcal{I}(\theta) = \mathbb{E} \left( \frac{\partial \ln \mathcal{L}_{\text{Exp}}(\vec{\xi}, \lambda, \theta)}{\partial \theta} \right)^2 = \mathbb{E} \left( \frac{n}{\lambda} \right)^2 = \left( \frac{n}{\lambda} \right)^2 = \frac{1}{D\theta_1^*}$$

Отже з нерівності Рао-Крамера оцінка є ефективною

Без надії сподіваючись, перевіримо оцінку на **асимптотичну нормальність**:

Центруємо і нормуємо нашу оцінку:

$$\frac{\theta_1^* - \mathbb{E}\theta_1^*}{\sqrt{D\theta_1^*}} = \frac{n(-\frac{\lambda}{n} + \min \xi_k - \theta)}{\lambda} \stackrel{(1)}{=} \frac{n}{\lambda} \left( -\frac{\lambda}{n} + \eta \right) \stackrel{(2)}{=} \psi - 1, \quad \eta \sim \text{Exp}\left(\frac{\lambda}{n}\right), \quad \psi \sim \text{Exp}(1)$$

Перехід (1):  $\min \xi_k \sim \text{Exp}\left(\frac{\lambda}{n}, \theta\right) = \eta + \theta$

Перехід (2):  $\frac{n}{\lambda} \eta \sim \text{Exp}(1)$  за властивістю експоненційного розподілу, яка доведена в додатку

$$\chi_{\psi-1}(t) = e^{-it} \cdot \chi_{\psi}(t) = e^{-it} \frac{1}{1-it} \not\rightarrow e^{-\frac{t^2}{2}}$$

Отже за теоремою Леві  $\frac{\theta_1^* - \mathbb{E}\theta_1^*}{\sqrt{D\theta_1^*}} \xrightarrow{F} N(0, 1)$ , тому оцінка не є асимптотично нормальною.

Остаточно отримали такі найкращі(ефективні) оцінки:

$$\lambda^{**} = \bar{\xi} - \theta, \quad \theta_1^* = \min_{1 \leq k \leq n} \xi_k - \frac{\lambda}{n}$$

Знайдемо значення цих точкових оцінок:

Оскільки  $\lambda^{**} = \lambda^{**}(\vec{\xi}, \theta)$ ,  $\theta_1^* = \theta_1^*(\vec{\xi}, \lambda)$ , а  $\theta$  і  $\lambda$  нам невідомі, ми не можемо знайти точні значення цих точкових оцінок. Проте ми можемо знайти приблизні значення замінюючи невідомі  $\theta, \lambda$  значеннями їх найкращих оцінок:

$$(\lambda^{**})_{\text{зн}} = \bar{x} - \theta \approx \bar{x} - (\theta_1^*)_{\text{зн}} = \bar{x} - \min x_k + \frac{\lambda}{n} \approx \bar{x} - \min x_k + \frac{(\lambda^{**})_{\text{зн}}}{n} \implies (\lambda^{**})_{\text{зн}} \approx \frac{n}{n-1}(\bar{x} - \min x_k)$$

$$(\theta_1^*)_{\text{зн}} = \min x_k - \frac{\lambda}{n} \approx \min x_k - \frac{(\lambda^{**})_{\text{зн}}}{n} = \min x_k - \frac{1}{n}(\bar{x} - \theta) \approx \min x_k - \frac{1}{n}(\bar{x} - (\theta_1^*)_{\text{зн}}) \implies (\theta_1^*)_{\text{зн}} \approx \frac{n}{n-1} \left( \min x_k - \frac{1}{n}\bar{x} \right)$$

В такому випадку  $(\lambda_{\text{зн}}^{**}) \approx \frac{100}{99}(5.4986 - 4.18) \approx 1.332$ ,  $(\theta_1^*)_{\text{зн}} \approx \frac{100}{99}(4.18 - \frac{1}{100} \cdot 5.4986) \approx 4.167$

Отже можемо висунути статистичну гіпотезу:  $H_0 : \xi \sim \text{Exp}(1.332, 4.167)$ . Перевіримо таку гіпотезу за критерієм згоди Пірсона.

## 8. Перевірити за допомогою критерію Пірсона гіпотезу про розподіл з рівнем значущості $\alpha = 0.05$

Перевіряємо гіпотезу  $H_0 : \xi \sim \text{Exp}(1.332, 4.167)$ , оскільки в якості параметрів розподілу ми використовуємо значення оцінок цих параметрів, то гіпотеза є складною.

Згідно гіпотези множина  $X$  можливих значень  $\xi$  – піввісь  $[4.167; +\infty)$ . Розіб'ємо цю множину на 9 таких підмножин, що попарно не перетинаються:  $[4.167; 4.18)$ ,  $[4.18; 4.82)$ ,  $[4.82; 5.45)$ ,  $[5.45; 6.09)$ ,  $[6.09; 6.73)$ ,  $[6.73; 7.36)$ ,  $[7.36; 8)$ ,  $[8; 13.09]$ ,  $(13.09; +\infty)$

$X_i$	$[4.167; 4.18)$	$[4.18; 4.82)$	$[4.82; 5.45)$	$[5.45; 6.09)$	$[6.09; 6.73)$	$[6.73; 7.36)$	$[7.36; 8)$	$[8; 13.09]$	$(13.09; +\infty)$
$p_i$	0.0096	0.3754	0.2331	0.1447	0.0899	0.0558	0.0346	0.0554	0.0015
$np_i$	0.96	37.54	23.31	14.47	8.99	5.58	3.46	5.54	0.15

Для того щоб для кожного  $X_i$  виконувалось  $np_i > 10$  об'єднаємо  $X_1$  з  $X_2$ ,  $X_5$  з  $X_6$ ,  $X_7$ ,  $X_8$ ,  $X_9$ . Отримаємо 4 підмножини  $X_i$

$X_i$	$[4.167; 4.82)$	$[4.82; 5.45)$	$[5.45; 6.09)$	$[6.09; +\infty)$
$p_i$	0.385	0.2331	0.1447	0.2372
$np_i$	38.5	23.31	14.47	23.72
$n_i$	36	21	20	23
$n_i - np_i$	-2.5	-2.31	5.53	-0.72

$$\eta = \frac{(n_1 - np_1)^2}{np_1} + \frac{(n_2 - np_2)^2}{np_2} + \frac{(n_3 - np_3)^2}{np_3} + \frac{(n_4 - np_4)^2}{np_4}$$

$$\eta_{\text{зн}} = \frac{(-2.5)^2}{38.5} + \frac{(-2.31)^2}{23.31} + \frac{5.53^2}{14.47} + \frac{(-0.72)^2}{23.72} \approx 2.5265$$

За теоремою Пірсона якщо  $H_0$  справджується, то  $\eta \xrightarrow{F} \chi_{4-2-1}^2 = \chi_1^2, n \rightarrow +\infty$

В такому випадку критична множина  $W = \{x : x \geq t_{0.05,1}\}$

А область прийняття гіпотези  $\bar{W} = \{x : x < t_{0.05,1}\}$ , де  $t_{0.05,1}$  – квантиль рівня 1-0.05 розподілу  $\chi_1^2$

З таблиці хі квадрат  $t_{0.05,1} \approx 3.84$

Як бачимо  $\eta_{\text{зн}} \approx 2.5265 < 3.84 \approx t_{0.05,1}$ , отже  $\eta_{\text{зн}} \in \bar{W}$

Отримали що на рівні значущості 0.05 дослідницькі дані не суперечать гіпотезі  $H_0 : \xi \sim \text{Exp}(1.332, 4.167)$

## 9. Знайти довірчий інтервал для параметрів гіпотетичного закону розподілу, взявши рівень надійності 0.95.

**Довірчий інтервал параметру  $\lambda$ :**

Ми довели що оцінка  $\lambda^{**} = \bar{\xi} - \theta$  – асимптотично нормальна, тому

$$\frac{\lambda^{**} - \mathbb{E}\lambda^{**}}{\sqrt{D\lambda^{**}}} = \frac{\sqrt{n}(\lambda^{**} - \lambda)}{\lambda} \approx N(0, 1)$$

Звідси

$$P\{|\lambda^{**} - \lambda| < \varepsilon\} = 0.95 \implies P\left\{\frac{\sqrt{n}|\lambda^{**} - \lambda|}{\lambda} < \frac{\varepsilon \cdot \sqrt{n}}{\lambda}\right\} = 0.95 \implies P\left\{\frac{\sqrt{n}|\lambda^{**} - \lambda|}{\lambda} < t_{0.475}\right\} = 0.95 = 2\Phi(t_{0.475})$$

З таблиці значень функції Лапласа  $t_{0.475} = 1.96$

Маємо

$$\frac{\sqrt{n}|\lambda^{**} - \lambda|}{\lambda} < t_{0.475} \implies \frac{10|\lambda^{**} - \lambda|}{\lambda} < 1.96 \implies |\lambda^{**} - \lambda| < 0.196\lambda \implies \begin{cases} -0.196\lambda < \lambda^{**} - \lambda \\ \lambda^{**} - \lambda < 0.196\lambda \end{cases} \implies \begin{cases} \lambda < \frac{\lambda^{**}}{0.804} \\ \lambda > \frac{\lambda^{**}}{1.196} \end{cases}$$

Отримали що з ймовірністю 0.95  $\lambda \in \left(\frac{(\lambda^{**})_{\text{зн}}}{1.196}; \frac{(\lambda^{**})_{\text{зн}}}{0.804}\right) \implies \lambda \in (1.11; 1.66)$

**Довірчий інтервал параметру  $\theta$ :**

Оскільки  $f_{\xi}(x) = 0$  при  $x < \theta$ , то  $P\{\theta \leq \min \xi_k\} = 1$

Звідси  $P\{\theta_1^* - \varepsilon < \theta \leq \min \xi_k\} = P\{\theta_1^* - \varepsilon < \theta\} = P\{\theta_1^* - \theta < \varepsilon\}$

При перевірці оцінки  $\theta_1^* = -\frac{\lambda}{n} + \min \xi_k$  на асимптотичну нормальність було доведено:

$$\frac{\theta_1^* - \mathbb{E}\theta_1^*}{\sqrt{D\theta_1^*}} = \frac{n(-\frac{\lambda}{n} + \min \xi_k - \theta)}{\lambda} = \psi, \quad \psi \sim \text{Exp}(1, -1)$$

Отримуємо:

$$P\{\theta_1^* - \varepsilon < \theta \leq \min \xi_k\} = P\{\theta_1^* - \theta < \varepsilon\} = P\left\{\frac{\theta_1^* - \mathbb{E}\theta_1^*}{\sqrt{D\theta_1^*}} < \frac{\varepsilon}{\sqrt{D\theta_1^*}}\right\} = P\left\{\underbrace{\frac{\theta_1^* - \mathbb{E}\theta_1^*}{\sqrt{D\theta_1^*}}}_{\sim \text{Exp}(1, -1)} < \frac{n\varepsilon}{\lambda}\right\} = 0.95 \implies$$

$$F_{\psi}\left(\frac{n\varepsilon}{\lambda}\right) = 0.95 \implies F_{\psi}\left(\frac{n\varepsilon}{(\lambda^{**})_{\text{зн}}}\right) \approx 0.95 \implies F_{\psi}\left(\frac{100\varepsilon}{1.332}\right) \approx 0.95 \implies 1 - e^{-\left(\frac{100\varepsilon}{1.332} + 1\right)} \approx 0.95 \implies$$

$$e^{-\left(\frac{100\varepsilon}{1.332} + 1\right)} \approx 0.05 \implies \frac{100\varepsilon}{1.332} + 1 \approx \ln\left(\frac{1}{0.05}\right) \implies \varepsilon \approx (\ln 20 - 1) \frac{1.332}{100} \approx 0.0266$$

Отже з ймовірністю 0.95  $\theta \in ((\theta_1^*)_{\text{зн}} - 0.0266; \min x_k] \implies \theta \in (4.14; 4.18]$

## 10. Висновки

Проаналізувавши дану конкретну реалізацію вибірки ми висунули припущення щодо закону розподілу ГС, знайшли найкращі точкові оцінки та, порахувавши їх приблизне значення, перевірили гіпотезу за допомогою критерію значущості Пірсона.

Ми прийшли до висновку що на заданому рівні значущості дослідні дані не суперечать гіпотезі, що ГС, за якою отримано нашу реалізацію вибірки, розподілена за показниковим законом з параметрами  $\lambda = 1.332, \theta = 4.167$ . Також ми знайшли наближені довірчі інтервали з рівнем надійності 0.95 і отримали що  $\lambda \in (1.11; 1.66), \theta \in (4.14; 4.18]$

## Додаток

Експоненційний розподіл із зсувом:  $\xi \sim \text{Exp}(\lambda, \theta)$

$$f_{\xi}(x) = \begin{cases} 0, & x < \theta \\ \frac{1}{\lambda} e^{-\frac{1}{\lambda}(x-\theta)}, & x \geq \theta \end{cases}$$

$$F_{\xi}(x) = \begin{cases} 0, & x < \theta \\ \frac{1}{\lambda} \int_{\theta}^x e^{-\frac{1}{\lambda}(t-\theta)} dt, & x \geq \theta \end{cases} = \langle u = \frac{1}{\lambda}(t-\theta) \rangle = \begin{cases} 0, & x < \theta \\ \int_0^{\frac{1}{\lambda}(x-\theta)} e^{-u} du, & x \geq \theta \end{cases} = \begin{cases} 0, & x < \theta \\ 1 - e^{-\frac{1}{\lambda}(x-\theta)}, & x \geq \theta \end{cases}$$

$$\mathcal{L}_{\text{Exp}}(\vec{x}, \lambda, \theta) = \prod_{k=1}^n f_{\xi}(x_k) = \lambda^{-n} \cdot e^{-\frac{1}{\lambda} \sum_{k=1}^n (x_k - \theta)}$$

$$\ln \mathcal{L}_{\text{Exp}}(\vec{x}, \lambda, \theta) = -n \ln \lambda - \frac{1}{\lambda} \sum_{k=1}^n (x_k - \theta) = -n \ln \lambda - \frac{1}{\lambda} \sum_{k=1}^n x_k + \frac{n\theta}{\lambda}$$

Розглянемо функцію  $\varphi(x) = x + \theta$ ,  $\varphi^{-1}(y) = y - \theta$

Візьмемо випадкову величину  $\eta \sim \text{Exp}(\lambda)$ , та знайдемо випадкову величину  $\widehat{\xi} = \varphi(\eta)$ , що є функцією випадкового аргументу

Оскільки  $\varphi(x)$  – монотонно зростаюча неперервна функція на  $\mathbb{R}$ ,  $\varphi^{-1}(x)$  – диференційовна на  $\mathbb{R}$ , то

$$f_{\widehat{\xi}}(y) = f_{\eta}(\varphi^{-1}(y)) |(\varphi^{-1}(y))'| = \begin{cases} 0, & y - \theta < 0 \\ \frac{1}{\lambda} e^{-\frac{1}{\lambda}(y-\theta)}, & y - \theta \geq 0 \end{cases} = \begin{cases} 0, & y < \theta \\ \frac{1}{\lambda} e^{-\frac{1}{\lambda}(y-\theta)}, & y \geq \theta \end{cases}$$

Можемо замітити що  $f_{\widehat{\xi}}(x) = f_{\xi}(x)$ , а щільність однозначно визначає розподіл, тому  $\xi = \widehat{\xi} = \eta + \theta$

Завдяки цьому знайдемо числові характеристики:

$$\mathbb{E}\xi = \mathbb{E}(\eta + \theta) = \mathbb{E}\eta + \mathbb{E}\theta = \lambda + \theta$$

$$D\xi = D(\eta + \theta) = D\eta = \lambda^2$$

$$\mathbb{E}(\xi - \mathbb{E}\xi)^3 = \mathbb{E}(\eta + \theta - \mathbb{E}(\eta + \theta))^3 = \mathbb{E}(\eta + \theta - \mathbb{E}\eta - \theta)^3 = \mathbb{E}(\eta - \mathbb{E}\eta)^3, D\xi = D\eta \implies As\xi = As\eta = 2$$

Оскільки щільність відмінна від нуля тільки на півосі  $[\theta; +\infty)$  при чому на ньому вона монотонно спадає, найбільшого значення щільність прийматиме в точці  $x_{max} = \theta_0$ , отже  $Mo\xi = \theta$

$$F_{\xi}(x_0) = \frac{1}{2} \implies \frac{1}{2} = 1 - e^{-\frac{1}{\lambda}(x_0-\theta)} \implies -\frac{1}{\lambda}(x_0 - \theta) = \ln \frac{1}{2} \implies x_0 = -\lambda \ln \frac{1}{2} + \theta \implies x_0 = \lambda \ln 2 + \theta$$

$$\text{Отже } Me\xi = \lambda \ln 2 + \theta$$

Знайдемо ще характеристичну функцію  $\xi$ :

$$\chi_{\xi}(t) = \chi_{\eta+\theta}(t) = e^{it\theta} \cdot \chi_{\eta}(t) = e^{it\theta} \cdot \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} - it}$$

Доведемо тепер таку властивість:  $\frac{1}{\lambda}\eta \sim \text{Exp}(1)$ , де  $\eta \sim \text{Exp}(\lambda)$ :

$$\chi_{\frac{1}{\lambda}\eta}(t) = \chi_{\eta}\left(\frac{1}{\lambda}t\right) = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} - \frac{1}{\lambda}it} = \frac{1}{1 - it} = \chi_{\zeta}(t), \quad \zeta \sim \text{Exp}(1)$$