Розрахункова робота

з математичної статистики

студентки групи КА-92

Катасанової Карини

Завдання:
1. Побудувати варіаційний (дискретний або інтервальний) ряд наданої вибірки.
2. Зробити графічне зображення вибірки.
3. Побудувати емпіричну функцію розподілу.
4. Обчислити значення вибіркової медіани, моди, асиметрії.
5. Знайти незміщену оцінку математичного сподівання та дисперсії.
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 (Пірсона) гіпотезу про розподіл з рівнем значущості $\alpha = 0{,}05$.
9. Знайти довірчий інтервал для параметрів гіпотетичного закону розподілу, взяв рівень надійності $\gamma = 0.95$.
10. Висновки.

Конкретна реалізація вибірки відповідно до варіанту:

6.71	5.66	3.06	4.25	3.40	6.45	3.46	2.57	6.55	8.50
3.63	7.22	7.61	2.72	3.91	2.98	7.05	11.79	7.51	4.44
13.59	6.61	4.53	12.28	9.50	4.26	4.52	5.55	11.34	4.21
4.02	4.34	2.67	2.95	8.86	4.87	5.76	7.68	2.75	2.62
2.86	5.16	2.88	6.62	5.03	3.30	5.47	7.80	7.80	5.43
3.05	3.99	4.10	2.85	3.28	4.70	6.67	2.85	4.71	6.94
7.17	3.02	2.97	10.96	2.75	3.03	3.61	15.50	6.24	2.89
5.57	5.75	4.86	10.36	14.49	15.38	2.80	3.86	4.67	4.58
8.64	3.65	3.65	4.53	3.63	11.68	3.14	6.07	6.57	3.22
6.12	3.37	3.90	6.63	2.88	9.18	2.96	4.86	3.18	6.59

Відсортуємо:

2.57	2.62	2.67	2.72	2.75	2.75	2.80	2.85	2.85	2.86
2.88	2.88	2.89	2.95	2.96	2.97	2.98	3.02	3.03	3.05
3.06	3.14	3.18	3.22	3.28	3.30	3.37	3.40	3.46	3.61
3.63	3.63	3.65	3.65	3.86	3.90	3.91	3.99	4.02	4.10
4.21	4.25	4.26	4.34	4.44	4.52	4.53	4.53	4.58	4.67
4.70	4.71	4.86	4.86	4.87	5.03	5.16	5.43	5.47	5.55
5.57	5.66	5.75	5.76	6.07	6.12	6.24	6.45	6.55	6.57
6.59	6.61	6.62	6.63	6.67	6.71	6.94	7.05	7.17	7.22
7.51	7.61	7.68	7.80	7.80	8.50	8.64	8.86	9.18	9.50
10.36	10.96	11.34	11.68	11.79	12.28	13.59	14.49	15.38	15.50

1. Побудувати варіаційний (дискретний або інтервальний) ряд наданої вибірки.

Оскільки майже всі значення конкретної реалізації вибірки є унікальними доцільно будувати саме інтервальний варіаційним ряд.
Для побудови інтервального варіаційного ряду потрібно поділити відрізок $[x_{min}\ ;\ x_{max}]$ на інтервали.
$x_{min} = 2.57$
$x_{max} = 15.50$
$n = 100$ - обсяг вибірки.
$R = x_{max} - x_{min} = 15.50 - 2.57 = 12.93$ - розмах вибірки.
Для поділу відрізка на інтервали скористаємося правилом Стерджеса. Знайдемо кількість інтервалів:

$$k = 1 + [3.322 \lg n] = 1 + [3.322 * 2] \approx 1 + 7 = 8$$

Проте для досить великого обсягу вибірки (100) 8 інтервалів може виявитися замало. Тому я поділю відрізок не на 8, а на 10 інтервалів. Шукаємо довжину інтервалу:

$$r = \frac{R}{k} = \frac{12.93}{10} = 1.293$$

Будуємо інтервальний варіаційний ряд:

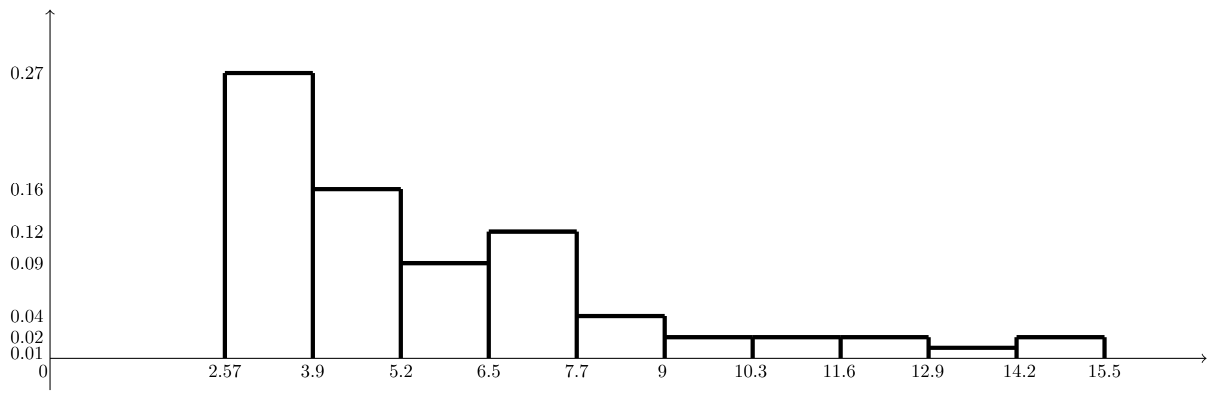
інтервал Δ_i	[2.57 ; 3.863)	[3.863 ; 5.156)	[5.156 ; 6.449)	[6.449 ; 7.742)	[7.742 ; 9.035)
частоти n_i	35	21	11	16	5
частості ω_i	0.35	0.21	0.11	0.16	0.05

інтервал Δ_i	[9.035 ; 10.328)	[10.328 ; 11.621)	[11.621 ; 12.914)	[12.914 ; 14.207)	[14.207 ; 15.50]
частоти n_i	2	3	3	1	3
частості ω_i	0.02	0.03	0.03	0.01	0.03

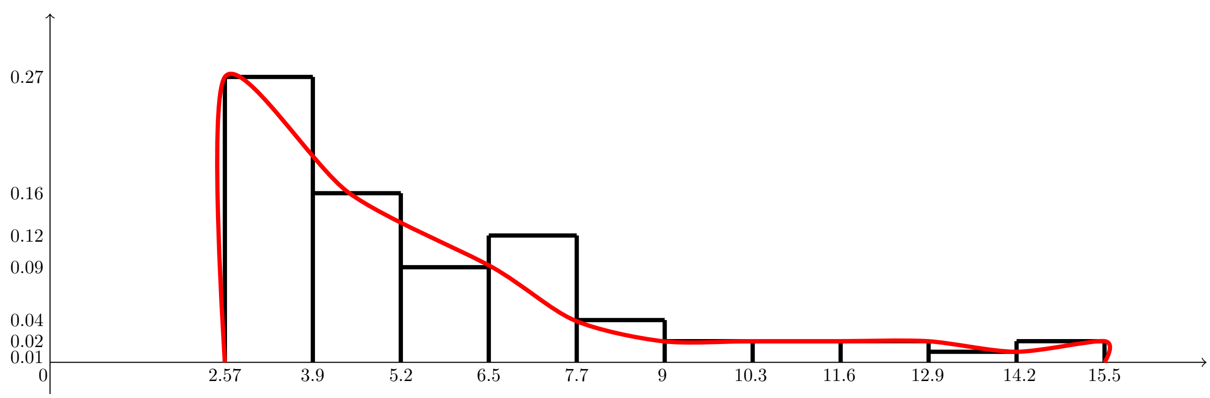
Деякі пояснення:
Частоти n_i - кількість варіант, які потрапили до i - того інтервалу.
Частості ω_i - відношення частості i - того інтервалу до обсягу вибірки.

2. Зробити графічне зображення вибірки.

Графічним зображенням (геометричною інтерпретацією) інтервального варіаційного ряду є гісторама, що складається з прямокутників з основами $\Delta_i = r = 1.293$ та висотами $h_i = \frac{\omega_i}{r}$. Для отриманого інтервального варіаційного ряду побудуємо наступну гістограму:



Зобразимо також "зглажену гістограму" (червоним):



Оскільки конкретна реалізація вибірки має багато унікальних значень, можна висунути припущення, що генеральна сукупність має неперервний розподіл. В такому випадку гістограма є "аналогом" щільності розподілу. Як видно з наведених зображень, вигляд гістограми подібний до графіка щільності експоненційного розподілу із зсувом, проте висувати гіпотезу про закон розподілу поки зарано.

3. Побудувати емпіричну функцію розподілу.

Для ІВР (інтервального варіаційного ряду) емпірична функція розподілу має вигляд:

$$F_n^*(x) = \begin{cases} 0, & x \le x_{min} \\ \omega_1^{\text{нак}} = \frac{n_1}{n}, & x_{min} < x \le t_1 \\ \omega_2^{\text{нак}} = \frac{n_1+n_2}{n}, & t_1 < x \le t_2 \\ ... \\ 1, & x > t_k \end{cases}$$

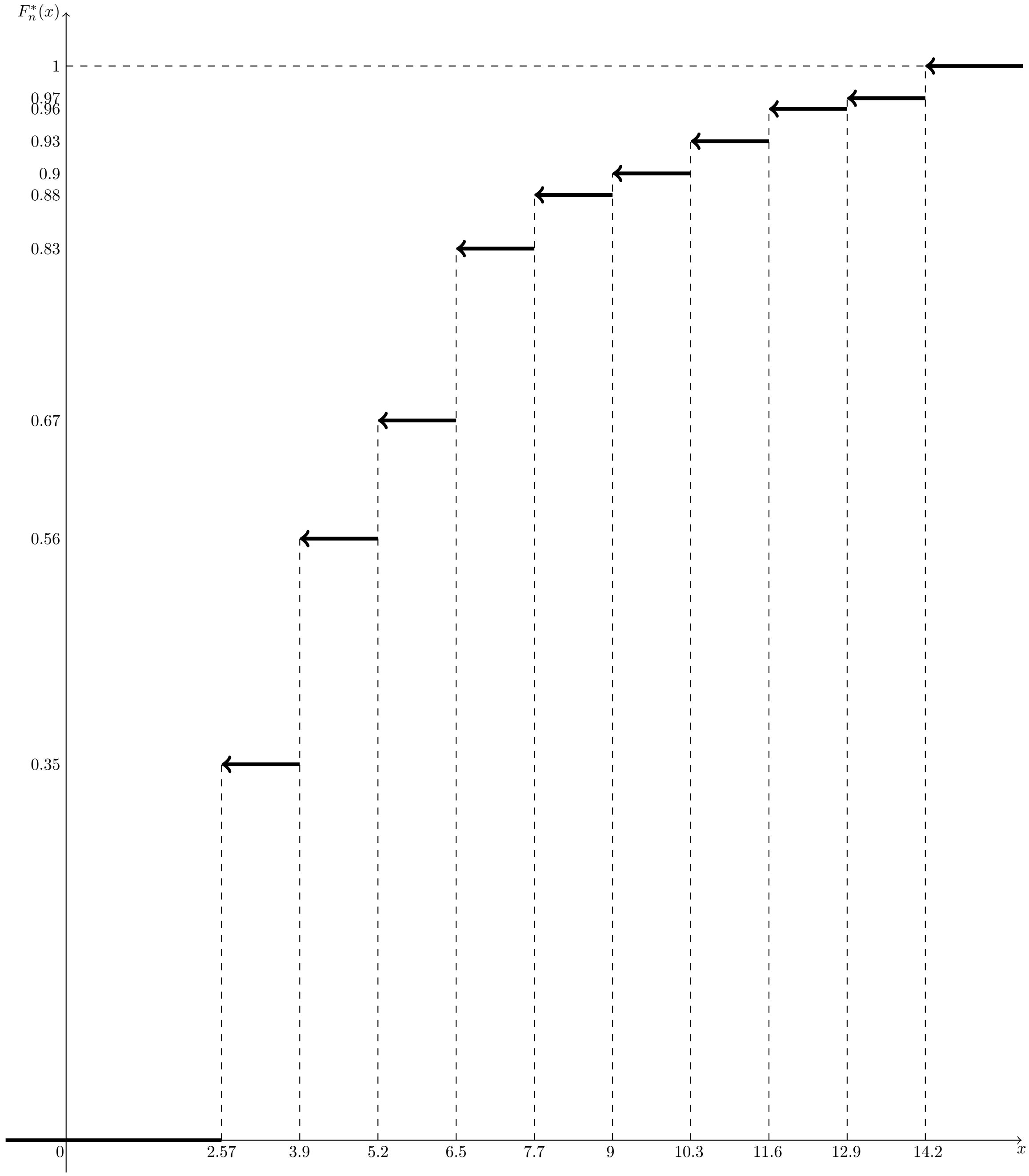
Тут, ω_i - накописені частоти, t_k - кінці інтервалів.
Для нашої конкретної реалізації маємо:

$$F_n^*(x) = \begin{cases} 0, & x \le 2.57 \\ \omega_1^{\text{нак}} = \frac{35}{100} = 0.35, & 2.57 < x \le 3.863 \\ \omega_2^{\text{нак}} = \frac{35+21}{100} = 0.56, & 3.863 < x \le 5.156 \\ \omega_3^{\text{нак}} = \frac{35+21+11}{100} = 0.67, & 5.156 < x \le 6.449 \\ \omega_4^{\text{нак}} = \frac{35+21+11+16}{100} = 0.83, & 6.449 < x \le 7.742 \\ \omega_5^{\text{нак}} = \frac{35+21+11+16+5}{100} = 0.88, & 7.742 < x \le 9.035 \\ \omega_6^{\text{нак}} = \frac{35+21+11+16+5+2}{100} = 0.9, & 9.035 < x \le 10.328 \\ \omega_7^{\text{нак}} = \frac{35+21+11+16+5+2+3}{100} = 0.93, & 10.328 < x \le 11.621 \\ \omega_8^{\text{нак}} = \frac{35+21+11+16+5+2+3+3}{100} = 0.96, & 11.621 < x \le 12.914 \\ \omega_9^{\text{нак}} = \frac{35+21+11+16+5+2+3+3+1}{100} = 0.97, & 12.914 < x \le 14.207 \\ \omega_{10}^{\text{нак}} = \frac{35+21+11+16+5+2+3+3+1+3}{100} = 1, & x > 14.207 \end{cases}$$

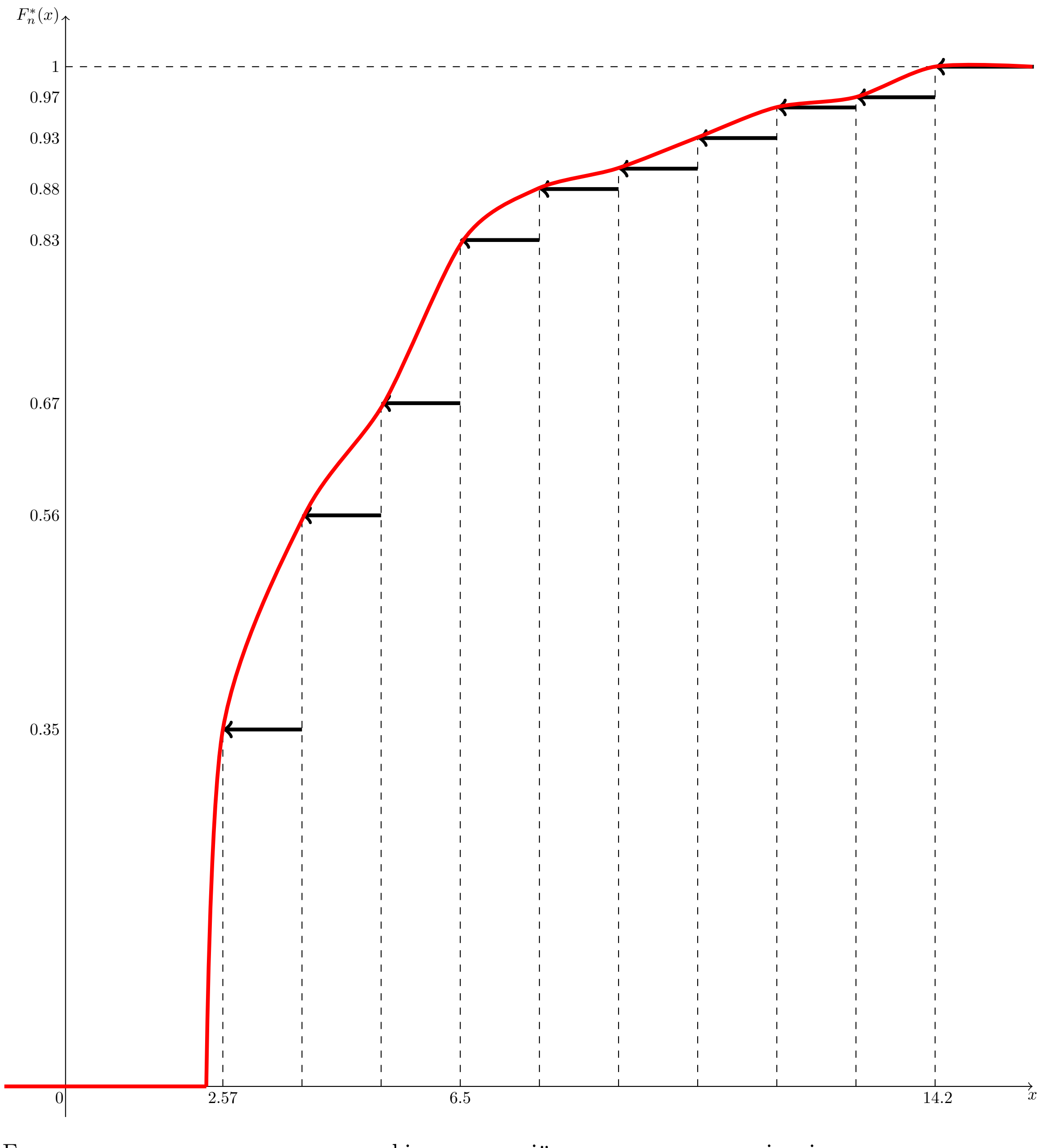
Остаточно маємо таку емпіричну функцію розподілу:

$$F_n^*(x) = \begin{cases} 0, & x \le 2.57 \\ 0.35, & 2.57 < x \le 3.863 \\ 0.56, & 3.863 < x \le 5.156 \\ 0.67, & 5.156 < x \le 6.449 \\ 0.83, & 6.449 < x \le 7.742 \\ 0.88, & 7.742 < x \le 9.035 \\ 0.9, & 9.035 < x \le 10.328 \\ 0.93, & 10.328 < x \le 11.621 \\ 0.96, & 11.621 < x \le 12.914 \\ 0.97, & 12.914 < x \le 14.207 \\ 1, & x > 14.207 \end{cases}$$

Побудуємо графік:


$F_n^*(x)$
1
0.97
0.96
0.93
0.9
0.88
0.83
0.67
0.56
0.35
0
2.57
3.9
5.2
6.5
7.7
9
10.3
11.6
12.9
14.2
x

Оскільки я підозрюю, що ГС(генеральна сукупність) має неперервний розподіл є сенс побудувати кумуляту (червоним):



Бачимо, що кумулята нагадує графік ексоненційного закону розподілу із зсувом.

4. Обчислити значення вибіркової медіани, моди, асиметрії.

У випадку ІВР вибіркова медіана обчислюється за формулою:

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

Тут:
y_i - початок інтервалу на якому накопичена частість перевищила 0.5.
h_i - довжина цього інтервалу, а n_i - частоти.
Маємо:

$$(Me^*\xi)_{\text{зн}} = 3.863 + \frac{1.293}{21} \cdot (50-35) = 3.863 + \frac{1.293 \cdot 5}{7} \approx 3.863 + 0.92357 = 4.78657$$

У випадку ІВР вибіркова мода обчислюється за формулою:

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

Тут:
y_i - початок інтервалу на який випадає найбільша частість.
h_i - довжина цього інтервалу, а n_i - частоти.
Маємо:

$$(Mo^*\xi)_{\text{зн}} = 2.57 + 1.293 \cdot \frac{35-0}{(35-0)+(35-21)} = 2.57 + 1.293 \cdot \frac{5}{7} \approx 2.57 + 0.92357 = 3.49357$$

Асиметрією називають таку статистику:

$$As^*\xi = \frac{\frac{1}{n}\sum_{k=1}^{n}(\xi_k - \overline{\xi})^3}{(D^*\xi)^{\frac{3}{2}}}$$

Як бачимо, для знахдження вибіркової асиметрії потрібні значення вибіркового середнього та вибіркової дисперсії, тож спочатку знайдемо їх.
У випадку ІВР значення вибіркового середнього обчислюється за формулою:

$$\overline{x} = \frac{1}{n} \sum_{k=1}^{r} x_k^* n_k$$

Тут за x_k^* приймається довільне значення k-того інтервалу. Візьмемо, наприклад, середину інтервалу. Для зручності обчислень випишемо середину кожного інтервалу:

$x_1^* = \frac{2.57+3.863}{2} = \mathbf{3.2165}$ $\quad$ $x_6^* = \frac{9.035+10.328}{2} = \mathbf{9.6815}$

$x_2^* = \frac{3.863+5.156}{2} = \mathbf{4.5095}$ $\quad$ $x_7^* = \frac{10.328+11.621}{2} = \mathbf{10.9745}$

$x_3^* = \frac{5.156+6.449}{2} = \mathbf{5.8025}$ $\quad$ $x_8^* = \frac{11.621+12.914}{2} = \mathbf{12.2675}$

$x_4^* = \frac{6.449+7.742}{2} = \mathbf{7.0955}$ $\quad$ $x_9^* = \frac{12.914+14.207}{2} = \mathbf{13.5605}$

$x_5^* = \frac{7.742+9.035}{2} = \mathbf{8.3885}$ $\quad$ $x_{10}^* = \frac{14.207+15.50}{2} = \mathbf{14.8535}$

Обчислюємо значення вибіркового середнього:

$$\overline{x} = \frac{3.2165 \cdot 35 + 4.5095 \cdot 21 + 5.8025 \cdot 11 + 7.0955 \cdot 16 + 8.3885 \cdot 5 + 9.6815 \cdot 2 + 10.9745 \cdot 3 + 12.2675 \cdot 3}{100} +$$

$$\frac{13.5605 \cdot 1 + 14.8535 \cdot 3}{100} = \frac{573.792}{100} = 5.73792$$

Вибірковою дисперсією називається наступна статистика (при невідомому математичному сподіванні):

$$D^*\xi = \frac{1}{n}\sum_{k=1}^{n}(\xi_k - \overline{\xi})^2$$

У випадку ІВР (за визначенням) маємо:

$$(D^*\xi)_{\text{зн}} = \frac{1}{n}\sum_{k=1}^{r}(x_k^* - \overline{x})^2 \cdot n_k$$, за x_k^* приймається довільне значення k-того інтервалу.

Обчислюємо:

$$(D^*\xi)_{\text{зн}} = \frac{(3.2165 - 5.73792)^2 \cdot 35 + (4.5095 - 5.73792)^2 \cdot 21 + (5.8025 - 5.73792)^2 \cdot 11}{100} +$$

$$+ \frac{(7.0955 - 5.73792)^2 \cdot 16 + (8.3885 - 5.73792)^2 \cdot 5 + (9.6815 - 5.73792)^2 \cdot 2}{100} +$$

$$+ \frac{(10.9745 - 5.73792)^2 \cdot 3 + (12.2675 - 5.73792)^2 \cdot 3 + (13.5605 - 5.73792)^2 \cdot 1 + (14.8535 - 5.73792)^2 \cdot 3}{100} \approx$$

$$\approx \frac{870.6154}{100} \approx 8.70615$$

Однак, вибіркова дисперсія є лише асимптотично незміщеною оцінкою дисперсії (буде доведено у наступному пункті), тому я вважаю доцільно скористатися значенням виправленої вибіркової дисперсії, яка є незміщеною оцінкою дисперсії. Виправленою вибірковою дисперсії називають наступну статистику:

$$D^{**}\xi = \frac{n}{n-1}D^*\xi$$

$$(D^{**}\xi)_{\text{зн}} = \frac{100}{99} \cdot 8.70615 \approx 8.7941$$

Тепер можемо обчислити значення асиметрії(позначатимемо її як виправлену):

$$(As^{**}\xi)_{\text{зн}} = \frac{\frac{1}{n}\sum_{k=1}^{r}(x_k^* - \overline{x})^3 \cdot n_k}{(D^{**}\xi)_{\text{зн}}^{\frac{3}{2}}}$$

Обчислимо чисельник (центральний емпіричний момент третього порядку m_3):

$$m_3 = \frac{(3.2165 - 5.73792)^3 \cdot 35 + (4.5095 - 5.73792)^3 \cdot 21 + (5.8025 - 5.73792)^3 \cdot 11}{100} +$$

$$+\frac{(7.0955-5.73792)^3\cdot 16+(8.3885-5.73792)^3\cdot 5+(9.6815-5.73792)^3\cdot 2}{100}+$$

$$+\frac{(10.9745-5.73792)^3\cdot 3+(12.2675-5.73792)^3\cdot 3+(13.5605-5.73792)^3\cdot 1+(14.8535-5.73792)^3\cdot 3}{100}\approx$$

$$\approx\frac{3672,7719}{100}\approx 36,7277$$

Нарешті,асиметрія:

$$(As^{**}\xi)_{\text{зн}}\approx\frac{36,7277}{26.0788}\approx 1.4083$$

5. Знайти незміщену оцінку математичного сподівання та дисперсії.

Точковою оцінкою невідомого параметра θ називається статистика $\theta^* = \theta^*(\xi)$, значення якої (знайдене за конкретною реалізацією вибірки) береться за наближене значення $(\theta)_{\text{зн}} = \theta^* x_1^*, ...x_n^* \approx \theta$.
Оцінка називається незміщеною, якщо $E\theta^* = \theta$.
Відомо, що незміщеною оцінкою математичного сподівання є вибіркове середнє.
Покажемо це:

$$E\overline{\xi} = E\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right) = \frac{1}{n}\sum_{k=1}^{n}E\xi_k = [\text{всі } \xi_k \text{ однаково розподілені як ГС}] =$$

$$= \frac{1}{n}\cdot n \cdot E\xi = E\xi$$

Значення вибіркового середнього було обчислено у попередньому пункті:

$$(E^*\xi)_{\text{зн}} = (\overline{\xi})_{\text{зн}} = 5.73792$$

Точне значення математичного сподівання нам невідомо. Тому незміщену оцінку для дисперсії шукатимемо замінюючи математичне сподівання на вибіркове середнє. Перевіримо, чи буде вибіркова дисперсія незміщеною оцінкою дисперсії:

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

$$= E\left(\frac{1}{n}\sum_{k=1}^{n}((\xi_k - E\xi)^2 - 2(\xi_k - E\xi)(\overline{\xi} - E\xi) + (\overline{\xi} - E\xi^2)\right) =$$

$$= E\left(\frac{1}{n}\sum_{k=1}^{n}(\xi_k - E\xi)^2 - 2(\overline{\xi} - E\xi)\cdot\frac{1}{n}\sum_{k=1}^{n}(\xi_k - E\xi) + (\overline{\xi} - E\xi)^2\right) =$$

$$= E\left(\frac{1}{n}\sum_{k=1}^{n}(\xi_k - E\xi)^2 - (\overline{\xi} - E\xi)^2\right) = D\xi - E(\overline{\xi} - E\xi)^2 = D\xi - D\overline{\xi} =$$

$$= D\xi - D\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right) = D\xi - \frac{1}{n^2}D\left(\sum_{k=1}^{n}\xi_k\right) = D\xi - \frac{1}{n}\cdot nD\xi = D\xi - \frac{D\xi}{n} = \frac{n-1}{n}D\xi$$

Як бачимо оцінка не є незміщеною, проте її можна виправити:

$$D^{**}\xi = \frac{n}{n-1}D^*\xi$$

Ця статистика - виправлена вибіркова дисперсія, яка і є незміщеною оцінкою дисперсії при невідомому математичному сподіванні. Її значення обчислили у попередньому пункті.

$$(D^{**}\xi)_{\text{зн}} = 8.7941$$

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

Провівши первинну обробку данних зауважимо:
1. Вибірка містить багато унікальних значень, отже можна припустити, що генеральна сукупність має неперервний розподіл (з канонічних це рівномірний, експоненційний та гауссівський).
2. З графічного зображення вибірки видно, що гістограма (що у випадку ІВР є, в певному сенсі, аналогом щільності) нагадує графік щільності експоненційного розподілу (зі зсувом).
3. Кумулята нагадує графік функції розподілу експоненційного розподілу (зі зсувом).
4. $(As^{**}\xi)_{\text{зн}} > 0$, якщо говорити про канонічні закони розподілу, то у рівномірного та нормального законів розподілу асиметрія нульова, а відмінна від нуля асиметрія лише у експоненційного розподілу(звісно, якщо говорити про канонічні).
5. Модальний інтервал чітко виражений (видно з гістограми). Крім того, значення моди лежить ближче до мінімального значення конкретної реалізації вибірки (що характерно для експоненційного закону), ніж до середини інтервалу [2,57 ; 15.50] (як би це було у рівномірного закону розподілу), і досить далеке від значення незміщеної оцінки математичного сподівання (як би це було у гауссівського закону розподілу).
6. Значення медіани також лежить ближче до лівого краю інтервалу, в той час як у нормального закону та у рівномірного медіана співпадає з математичним сподіванням та знаходиться посередині відповідно.
Отже, висувається основна гіпотеза:
$H_0 : \xi \sim Exp\left(\frac{1}{\lambda}, a\right)$- генеральна сукупність розподілена за експоненційним законом зі зсувом. Для зручності подальших досліджень першим параметром візьмемо $\frac{1}{\lambda}$.
Це статистична непараметрична гіпотеза. Найкращі точкові оцінки параметрів будуть обчислені пізніше.
Також висунемо альтернативну гіпотезу:
$H_1 : \xi \nsim Exp\left(\frac{1}{\lambda}, a\right)$- генеральна сукупність розподілена за іншим законом.

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

Вважаємно, що $\xi \sim Exp(\frac{1}{\lambda}, a)$.

1. Метод моментів

Відомо, що $E\xi = \frac{1}{\frac{1}{\lambda}} + a$. Але щоб виразити через моменти обидва параметри необхідне ще одне рівняння: $D\xi = \frac{1}{\frac{1}{\lambda^2}}$. Маємо систему:

$$\begin{cases} \lambda^* + a^* = \overline{\xi} \\ \lambda^* = \sqrt{D^{**}\xi} \end{cases} \quad \Rightarrow \quad \begin{cases} a^*_{\text{мм}} = \overline{\xi} - \sqrt{D^{**}\xi} \\ \lambda^*_{\text{мм}} = \sqrt{D^{**}\xi} \end{cases}$$

2. Метод максимальної правдоподібності

Щільність розподілу експоненційного закону зі зсувом має вигляд:

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

Можемо знайти функцію правдоподібності:

$$\mathcal{L}_{Exp}(\overrightarrow{x}, \lambda, a) = \prod_{k=1}^{n} \frac{1}{\lambda} e^{-\frac{1}{\lambda}(x_k - a)} = \lambda^{-n} e^{-\frac{1}{\lambda}\left(\sum_{k=1}^{n} x_k - na\right)}$$

Логарифмуємо:

$$\ln \mathcal{L}_{Exp}(\overrightarrow{x}, \lambda, a) = -n \ln \lambda - \frac{1}{\lambda} \left(\sum_{k=1}^{n} x_k - na \right)$$

$$\begin{cases} \frac{\partial \ln \mathcal{L}_{Exp}(\overrightarrow{x}, \lambda, a)}{\partial \lambda} = -\frac{n}{\lambda} + \frac{1}{\lambda^2} \left(\sum_{k=1}^{n} x_k - na \right) \\ \\ \frac{\partial \ln \mathcal{L}_{Exp}(\overrightarrow{x}, \lambda, a)}{\partial a} = \frac{n}{\lambda} \end{cases}$$

З першого рівняння системи знайдемо $\lambda_{\text{кр}}$:

$$-\frac{n}{\lambda} + \frac{1}{\lambda^2} \left(\sum_{k=1}^{n} x_k - na \right) = 0 \quad \Rightarrow \quad \frac{1}{\lambda} \left(\frac{1}{\lambda} \left(\sum_{k=1}^{n} x_k - na \right) - n \right) = 0 \quad \Rightarrow \quad \frac{1}{\lambda} \left(\sum_{k=1}^{n} x_k - na \right) - n = 0 \quad \Rightarrow$$

$$\Rightarrow \quad \frac{1}{\lambda} \left(\sum_{k=1}^{n} x_k - na \right) = n \quad \Rightarrow \quad \lambda n = \sum_{k=1}^{n} x_k - na \quad \Rightarrow \quad \lambda_{\text{кр}} = \frac{\sum_{k=1}^{n} x_k - na}{n} = \overline{x} - a$$

Шукаємо другу похідну:

$$\frac{\partial^2 \ln \mathcal{L}_{Exp}(\overrightarrow{x}, \lambda, a)}{\partial \lambda^2} = \frac{n}{\lambda^2} - \frac{2}{\lambda^3} \left(\sum_{k=1}^{n} x_k - na \right)$$

Підставимо знайдену критичну точку:

$$\frac{\partial^2 \ln \mathcal{L}_{Exp}(\overrightarrow{x}, \lambda_{\text{кр}}, a)}{\partial \lambda^2} = \frac{1}{\lambda_{\text{кр}}^2} \left(n - \frac{2}{\lambda_{\text{кр}}} \left(\sum_{k=1}^{n} x_k - na \right) \right) = \frac{n}{(\overline{x} - a)^2} \left(n - \frac{2}{\overline{x} - a} (n\overline{x} - na) \right) =$$

$$= \frac{n}{(\overline{x} - a)^2} \cdot (-n) < 0$$

Отже, робимо висновок, що отримали оцінку $\lambda^*_{\text{ммп}} = \overline{\xi} - a$.
З другого рівняння системи ми маємо: $\frac{n}{\lambda}$ - цей вираз завжди додатній, отже робимо висновок, що $\ln \mathcal{L}_{Exp}(\overrightarrow{x}, \lambda, a)$ монотонно зростає відносно a.Відомо, що $x_k > a, \quad \forall \quad k = 1, ..., n$, отже найбільше значення яке може прийняти параметр $a = \min\limits_{1 \le k \le n} x_k$. Отримали оцінку $a^*_{\text{ммп}} = \min\limits_{1 \le k \le n} \xi_k$.
Очевидно, що оцінки отримані методом максимальної правдоподібності буде досліджувати простіше. Оскільки для дослідження параметра λ нам потрібно вважати параметр a відомим, дослідимо властивості його оцінки в першу чергу.

ПАРАМЕТР a.

Статистика $a^*_{\text{ммп}} = \min\limits_{1 \le k \le n} \xi_k$ є випадковою величиною. Знайдемо її щільність розподілу (беручи до уваги, що усі ξ_k розподілені однаково як ГС):

$$f_{a^*_{\text{ммп}}}(x) = n(F_\xi(x))^{n-1} f_\xi(x) = n(1-(1-e^{-\frac{1}{\lambda}(x-a)})^{n-1}) \cdot \frac{1}{\lambda} e^{-\frac{1}{\lambda}(x-a)} = \frac{n}{\lambda} e^{-\frac{n}{\lambda}(x-a)}, \ x \ge a$$

Отже, $a^*_{\text{ммп}} \sim Exp\left(\frac{n}{\lambda}, a\right)$. Тепер перейдемо до досліджень ($a^*_{\text{ммп}}$ позначатимемо як a^*):
1. Незміщеність

$$Ea^* = \frac{\lambda}{n} + a$$

$$\lim_{n \to \infty} \left(\frac{\lambda}{n} + a \right) = a$$

Маємо асимптотично незміщену оцінку.
2. Консистентність
За означенням точкова оцінка θ^* невідомого параметра θ називається консистентною, якщо

$$\forall \, \varepsilon > 0: \lim_{n \to \infty} \mathbb{P}\{|\theta^*_n - \theta| \ge \varepsilon\} = 0 \iff \theta^*_n \xrightarrow{P} \theta, \ n \to \infty$$

Оскільки маємо асимптотично незміщену оцінку можемо скористатися достатньою умовою консистентності:

Якщо θ^*_n незміщена чи асомптотично незміщена оцінка та $D\theta^*_n \to 0, \ n \to \infty$, то ця оцінка є консистентною.

Маємо:

$$Da^* = \frac{\lambda^2}{n^2} \rightarrow 0, \; n \rightarrow \infty.$$

Маємо консистентну оцінку.

3. Ефективність

Поняття ефективності вводиться лише для незміщених оцінок. Знайдемо коефіцієнт ефективності оцінки a^*.

$$\frac{1}{D\theta_n^* \cdot I(\theta)} = e(\theta^*),$$

де $I(\theta) = E\left(\frac{\partial \ln L(\overrightarrow{\xi},\theta)}{\partial \theta}\right)^2$ - кількість інформації за Фішером.

Дисперсію знайшли під час дослідження оцінки на консистентність, а кількість інформації за Фішером:

$$I(a) = E\left(\frac{n}{\lambda}\right)^2 = \frac{n^2}{\lambda^2}$$

Отже, маємо:

$$e(a^*) = \frac{1}{\frac{\lambda^2}{n^2} \cdot \frac{n^2}{\lambda^2}} = 1$$

Коефіцієнт ефективності оцінки $e(a^*) = 1$.

Отже, робимо висновок, що оцінка a^* є асимптотично незміщеною та консистентною, коефіцієнт ефективності $e(a^*) = 1$.

ПАРАМЕТР λ.

Будемо досліджувати оцінку $\lambda^*_{\text{ммп}} = \overline{\xi} - a$, але оскільки параметр a невідомий, замінимо його на досліджену оцінку a^*, отже досліджуємо властивості оцінки $\lambda^*_{\text{ммп}} = \overline{\xi} - \min\limits_{1 \le k \le n} \xi_k$. При подальших дослідженнях позначатимемо $\lambda^*_{\text{ммп}}$ як λ^*.

1. Незміщеність

$$E\lambda^* = E\left(\overline{\xi} - \min_{1 \le k \le n} \xi_k\right) = E\overline{\xi} - E(\min \xi_k) = E\left(\frac{1}{n}\left(\sum_{k=1}^{n} \xi_k\right)\right) - \frac{\lambda}{n} - a = E\xi - \frac{\lambda}{n} - a = \lambda + a - \frac{\lambda}{n} - a =$$

$$= \lambda - \frac{\lambda}{n} = \frac{n-1}{n}\lambda$$

$$\lim_{n \to \infty}\left(\frac{n-1}{n}\lambda\right) = \lambda$$

Отже, оцінка λ^* є асимптотично незміщеною, однак її можна виправити.

$$\lambda^{**} = \frac{n}{n-1}\lambda^* = \frac{n}{n-1}\left(\overline{\xi} - \min_{1 \le k \le n} \xi_k\right)$$

λ^{**} **- незміщена оцінка параметра** λ, тому надалі досліджуватимемо саме її властивості.
2. Консистентність
Оскільки маємо незміщену оцінку, можемо скористатися достатньою умовою консистентності.

$$D\lambda^{**} = D\left(\frac{n}{n-1}\left(\overline{\xi} - \min_{1\le k\le n}\xi_k\right)\right) = \frac{n^2}{(n-1)^2}D\left(\overline{\xi} - \min_{1\le k\le n}\xi_k\right) =$$

$$= \frac{n^2}{(n-1)^2}\left(D\overline{\xi} - 2cov(\overline{\xi}, \min_{1\le k\le n}\xi_k) + D\min_{1\le k\le n}\xi_k\right) = \left[|cov(\xi,\eta)| \le \sqrt{D\xi\cdot D\eta}\right] \le$$

$$\le \frac{n^2}{(n-1)^2}\left(D\overline{\xi} + 2\sqrt{D\overline{\xi}\cdot D\min_{1\le k\le n}\xi_k} + D\min_{1\le k\le n}\xi_k\right) =$$

$$= \frac{n^2}{(n-1)^2}\left(\sqrt{D\overline{\xi}} + \sqrt{D\min_{1\le k\le n}\xi_k}\right)^2 = \left[\begin{array}{l} D\overline{\xi} = \frac{1}{n^2}D\left(\sum_{k=1}^n \xi_k\right) = \frac{1}{n}D\xi = \frac{\lambda^2}{n} \\ \\ D\min\limits_{1\le k\le n}\xi_k = \frac{\lambda^2}{n^2} \end{array}\right] =$$

$$= \frac{n^2}{(n-1)^2}\left(\sqrt{\frac{\lambda^2}{n}} + \sqrt{\frac{\lambda^2}{n^2}}\right)^2 = \frac{n^2}{(n-1)^2}\left(\frac{\lambda}{\sqrt{n}} + \frac{\lambda}{n}\right)^2 \ \rightarrow\ 0,\ n\ \rightarrow\infty$$

Отже, оцінка λ^{**} - консистентна.
3. Ефективність
Для дослідження скористаємося критерієм ефективності незміщеної оцінки (необхідна і достатня умова ефективності оцінки):

Незміщена оцінка θ^* є оптимальною тоді і тільки тоді, коли $\dfrac{\partial \ln \mathcal{L}(\overrightarrow{\xi},\theta)}{\partial\theta} = C(n,\theta)\cdot\left(\theta^*\left(\overrightarrow{\xi}-\theta\right)\right)$.

Тут $C(n,\theta)$ - деяке значення, яке залежить тільки від n та θ. Маємо:

$$\frac{\partial \ln \mathcal{L}_{Exp}(\overrightarrow{x},\lambda,a)}{\partial\lambda} = -\frac{n}{\lambda} + \frac{1}{\lambda^2}\left(\sum_{k=1}^n x_k - na\right)$$

Даний вираз не можна представити у вигляді $C(n,\theta)\cdot\left(\theta^*\left(\overrightarrow{\xi}-\theta\right)\right)$, отже оцінка не є ефективною.
Спробуємо перевірити λ^{**} на асимптотичну ефективність.
Оцінка є асимптотично ефективною якщо $D\theta^*\cdot I(\theta)\ \rightarrow\ 1,\ n\ \rightarrow\ \infty$.
Знайдемо кількість інформації за Фішером:

$$I(\lambda) = E\left(-\frac{n}{\lambda} + \frac{1}{\lambda^2}\left(\sum_{k=1}^n \xi_k - na\right)\right)^2 = E\left(-\frac{n}{\lambda} + \frac{n(\overline{\xi}-a)}{\lambda^2}\right)^2 =$$

$$= E\left(\frac{n^2}{\lambda^2} - \frac{2n^2(\overline{\xi}-a)}{\lambda^3} + \frac{n^2(\overline{\xi}^2 - 2a\overline{\xi} + a^2)}{\lambda^4}\right) = \frac{n^2}{\lambda^4}E\left(\lambda^2 - 2\lambda(\overline{\xi}-a) + \overline{\xi}^2 - 2a\overline{\xi} + a^2\right) =$$

$$= \frac{n^2}{\lambda^4}\left(E\lambda^2 - E(2\lambda(\overline{\xi}-a) + E\overline{\xi}^2 - E2a\overline{\xi} + Ea^2\right) =$$

$$= \frac{n^2}{\lambda^4}\left(\lambda^2 + a^2 - 2\lambda E(\overline{\xi}-a) + E\overline{\xi}^2 - 2aE\overline{\xi}\right) = \left[\begin{array}{l} E\overline{\xi} = E\xi = \lambda + a \\ E\overline{\xi}^2 = D\overline{\xi} + (E\overline{\xi})^2 = \frac{\lambda^2}{n} + (\lambda+a)^2 \end{array}\right] =$$

$$= \frac{n^2}{\lambda^4}\left(\lambda^2 + a^2 - 2\lambda E\overline{\xi} + 2\lambda Ea + \frac{\lambda^2}{n} + (\lambda+a)^2 - 2a(\lambda+a)\right) =$$

$$= \frac{n^2}{\lambda^4}\left(\lambda^2 + a^2 - 2\lambda(\lambda+a) + 2\lambda a + \frac{\lambda^2}{n} + \lambda^2 + 2a\lambda + a^2 - 2\lambda a - 2a^2\right) =$$

$$= \frac{n^2}{\lambda^4}\left(2\lambda^2 - 2\lambda^2 - 2\lambda a + 2\lambda a + \frac{\lambda^2}{n}\right) = \frac{n}{\lambda^2}$$

Дисперсію оцінки ми оцінили під час перевірки на консистентність, дещо спростимо отриманий вираз:

$$D\lambda_1^{**} = \frac{n^2}{(n-1)^2}\left(\frac{\lambda}{\sqrt{n}} + \frac{\lambda}{n}\right)^2 = \frac{n^2}{(n-1)^2}\left(\frac{\lambda^2}{n} + \frac{2\lambda^2}{n\sqrt{n}} + \frac{\lambda^2}{n^2}\right) = \frac{n^2}{(n-1)^2}\cdot\frac{\lambda^2 n + 2\lambda^2\sqrt{n} + \lambda^2}{n^2} =$$

$$= \frac{\lambda^2(n + 2\sqrt{n} + 1)}{(n-1)^2} = \frac{\lambda^2(\sqrt{n}+1)^2}{(n-1)^2}$$

Маємо:

$$D\lambda_1^{**} \cdot I(\lambda) = \frac{\lambda^2(\sqrt{n}+1)^2}{(n-1)^2} \cdot \frac{n}{\lambda^2} = \left(\frac{n+\sqrt{n}}{n-1}\right)^2$$

$$\lim_{n\to\infty}\left(\frac{n+\sqrt{n}}{n-1}\right)^2 = \left(\lim_{n\to\infty}\frac{n+\sqrt{n}}{n-1}\right)^2 = \left(\lim_{n\to\infty}\frac{\sqrt{n}(\sqrt{n}+1)}{n-1}\right)^2 = \left(\lim_{n\to\infty}\frac{1+\frac{1}{\sqrt{n}}}{1-\frac{1}{n}}\right)^2 = 1$$

Оскільки $D\lambda^{**} \le D\lambda_1^{**}$, то $\lim_{n\to\infty} I(\lambda)\cdot D\lambda^{**} \le \lim_{n\to\infty} I(\lambda)\cdot D\lambda_1^{**} = 1$.
Крім того, згідно з підручником (1), для λ^{**}, як для незміщеної оцінки, виконується нерівність Крамера-Рао:

$$I(\lambda)\cdot D\lambda^{**} \ge 1$$

Тоді, за лемою про двох поліцейських маємо:

$$\lim_{n\to\infty} I(\lambda)\cdot D\lambda^{**} = 1$$

Також, порахуємо коефіцієнт ефективності:

$$e(\lambda^{**}) = \frac{1}{D\lambda_n^{**}\cdot I(\lambda)} \ \to\ 1,\ n\ \to\ \infty.$$

Оцінка λ^{**} асимптотично ефективна, з коефіцієнтом ефективності $\mathbf{e}(\lambda^{**}) \to 1,\ n \to \infty$.
Отже, робимо висновок, що оцінка λ^{**} є незміщеною, консистентною, асимптотично ефективною.
Коефіцієнт ефективності $e\left(\lambda^{**}\right) \to 1,\ n \to \infty$.
В подальшому користуватимемося дослідженими оцінками, якщо не вказано іншого.

8. Перевірити за допомогою критерію (Пірсона) гіпотезу про розподіл з рівнем значущості $\alpha = 0{,}05$.

У попередньому пункті були знайдені точкові оцінки парметрів розподілу експоненційного закону зі зсувом. Приймемо значення оцінок за наближене значення параметрів:

$$\lambda = \lambda^{**}_{\text{зн}} = \frac{n}{n-1}(\overline{x} - \min_{1\le k\le n} x_k) = \frac{100}{99}(5.73792 - 2.57) \approx 3.1999$$

$$a = a^{*}_{\text{зн}} = \min_{1\le k\le n} x_k = 2.57$$

Отже, перевірятимемо гіпотезу $H_0 : \xi \sim Exp(\frac{1}{3.1999}, 2.57)$. X = [2.57;+$\infty$] розіб'ємо на 10 підмножин X_i, $i = \overline{1,10}$, що попарно не перетинаються. Використамо інтервали, отримані під час побудови ІВР, при цьому розширивши останній на нескінченність:

$$X_1 = [2.57\ ;\ 3.863),\quad X_2 = [3,863\ ;\ 5.156),\quad X_3 = [5.156\ ;\ 6.449),\quad X_4 = [6.449\ ;\ 7.742),$$

$$X_5 = [7.742\ ;\ 9.035),\quad X_6 = [9.035\ ;\ 10.328),\quad X_7 = [10.328\ ;\ 11.621),\quad X_8 = [11.621\ ;\ 12.914),$$

$$X_9 = [12.914\ ;\ 14.207),\quad X_{10} = [14.207\ ;\ +\infty)$$

Обчислимо ймовірності $p_i = P\{\xi \in X_i/H_0\}$. Для зручності, введемо функцію розподілу експоненційного закону зі зсувом (підствивши значення параметрів):

$$F_\xi(x) = \begin{cases} 1 - e^{-\frac{(x-2.57)}{3.1999}}, & x > 2.57 \\ 0, & x \le 2.57 \end{cases}$$

Далі, позначатимемо функцію розподілу як F. Маємо:

$$p_1 = P\left(\xi \in [2.57\ ;\ 3.863)\right) = F(3.863) - F(2.57) = 1 - e^{-\frac{1.293}{3.1999}}$$

$$p_2 = P\left(\xi \in [3,863\ ;\ 5.156)\right) = F(5.156) - F(3,863) = e^{-\frac{1.293}{3.1999}} - e^{-\frac{2.586}{3.1999}}$$

$$p_3 = P\left(\xi \in [5.156\ ;\ 6.449)\right) = F(6.449) - F(5.156) = e^{-\frac{2.586}{3.1999}} - e^{-\frac{3.879}{3.1999}}$$

$$p_4 = P\left(\xi \in [6.449\ ;\ 7.742)\right) = F(7.742) - F(6.449) = e^{-\frac{3.879}{3.1999}} - e^{-\frac{5.182}{3.1999}}$$

$$p_5 = P\left(\xi \in [7.742\ ;\ 9.035)\right) = F(9.035) - F(7.742) = e^{-\frac{5.182}{3.1999}} - e^{-\frac{6.465}{3.1999}}$$

$$p_6 = P\left(\xi \in [9.035\ ;\ 10.328)\right) = F(10.328) - F(9.035) = e^{-\frac{6.465}{3.1999}} - e^{-\frac{7.758}{3.1999}}$$

$$p_7 = P\left(\xi \in [10.328\ ;\ 11.621)\right) = F(11.621) - F(10.328) = e^{-\frac{7.758}{3.1999}} - e^{-\frac{9.051}{3.1999}}$$

$$p_8 = P\left(\xi \in [11.621\ ;\ 12.914)\right) = F(12.914) - F(11.621) = e^{-\frac{9.051}{3.1999}} - e^{-\frac{10.344}{3.1999}}$$

$$p_9 = P\left(\xi \in [12.914\ ;\ 14.207)\right) = F(12.914) - F(12.914) = e^{-\frac{10.344}{3.1999}} - e^{-\frac{11.637}{3.1999}}$$

$$p_{10} = P\left(\xi \in [14.207\ ;\ +\infty)\right) = 1 - F(14.207) = e^{-\frac{11.637}{3.1999}}$$

Перевіримо, що $\sum_{k=1}^{10} p_k = 1$:

$$\sum_{k=1}^{10} p_k = 1-e^{-\frac{1.293}{3.1999}}+e^{-\frac{1.293}{3.1999}}-e^{-\frac{2.586}{3.1999}}+e^{-\frac{2.586}{3.1999}}-e^{-\frac{3.879}{3.1999}}+e^{-\frac{3.879}{3.1999}}-e^{-\frac{5.182}{3.1999}}+$$

$$e^{-\frac{5.182}{3.1999}}-e^{-\frac{6.465}{3.1999}}+e^{-\frac{6.465}{3.1999}}-e^{-\frac{7.758}{3.1999}}+e^{-\frac{7.758}{3.1999}}-e^{-\frac{9.051}{3.1999}}+e^{-\frac{9.051}{3.1999}}-e^{-\frac{10.344}{3.1999}}+$$
$$+e^{-\frac{10.344}{3.1999}}-e^{-\frac{11.637}{3.1999}}+e^{-\frac{11.637}{3.1999}}=1$$

Оскільки $r < 20$, потрібно щоб виконувалась умова $np_i \geq 10, \; i = \overline{1, r}$, де $n = 100$ - обсяг вибірки. Перевіримо:

$$np_1 = 100\left(1-e^{-\frac{1.293}{3.1999}}\right) \approx 32.968, \quad np_2 = 100\left(e^{-\frac{1.293}{3.1999}}-e^{-\frac{2.586}{3.1999}}\right) \approx 22.546$$

$$np_3 = 100\left(e^{-\frac{2.586}{3.1999}}-e^{-\frac{3.879}{3.1999}}\right) \approx 14.666, \quad np_4 = 100\left(e^{-\frac{3.879}{3.1999}}-e^{-\frac{5.182}{3.1999}}\right) \approx 10.029$$

$$np_5 = 100\left(e^{-\frac{5.182}{3.1999}}-e^{-\frac{6.465}{3.1999}}\right) \approx 6.5, \quad np_6 = 100\left(e^{-\frac{6.465}{3.1999}}-e^{-\frac{7.758}{3.1999}}\right) \approx 4.37$$

$$np_7 = 100\left(e^{-\frac{7.758}{3.1999}}-e^{-\frac{9.051}{3.1999}}\right) \approx 2.93, \quad np_8 = 100\left(e^{-\frac{9.051}{3.1999}}-e^{-\frac{10.344}{3.1999}}\right) \approx 2$$

$$np_9 = 100\left(e^{-\frac{10.344}{3.1999}}-e^{-\frac{11.637}{3.1999}}\right) \approx 1.33, \quad np_{10} = 100\cdot e^{-\frac{11.637}{3.1999}} \approx 2.625$$

Бачимо, що $np_{10} < np_9 < np_8 < np_7 < np_6 < np_5 < 10$. Для того щоб умова виконувалася об'єднаємо множини $X_5, \; X_6, \; X_7, X_8, \; X_9, \; X_{10}$. Позначимо отримані множини так:

$$X_1^{\sim} = X_1, \quad X_2^{\sim} = X_2, \quad X_3^{\sim} = X_3, \quad , X_4^{\sim} = X_4$$

$$X_5^{\sim} = X_5 \cup X_6 \cup X_7 \cup X_8 \cup X_9 \cup X_{10}$$

Для зручності подальших досліджень зведемо таблицю:

X_i	[2.57 ; 3.863)	[3,863 ; 5.156)	[5.156 ; 6.449)	[6.449 ; 7.742)	[7.742 ; $+\infty$)
n_i	35	21	11	16	17
p_i	0.3297	0.2255	0.1467	0.1003	0.1978
np_i	32.97	22.55	14.67	10.03	19.78

За теоремою Пірсона, гіпотеза H_0 справджується, якщо:

$$\eta = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i} \xrightarrow{F} \chi^2_{r-s-1}, \; n \;\to\; \infty$$

Обчислимо значення цієї статистики:

$$\eta_{\text{зн}} = \frac{(35-32.97)^2}{32.97}+\frac{(21-22.55)^2}{22.55}+\frac{(11-14.67)^2}{14.67}+\frac{(16-10.03)^2}{10.03}+\frac{(17-19.78)^2}{19.78} \approx 5.10$$

За таблицею розподілу χ^2 знаходимо $t_{\text{кр}} = t_{\alpha, r-s-1}$, де r - кількість інтервалів, s- кількість невідомих параметрів гіпотетичного закону розподілу. В нашому випадку: $r = 5$, $s = 2$. Отже:

$$t_{\text{кр}} = t_{0.05,2} = 5.99$$

Бачимо, що $\eta_{\text{зн}} = 5.10 < t_{\text{кр}} = 5.99$.

Отже, на рівні значущості $\alpha = 0.05$, дані не суперечать висунутій гіпотезі H_0.

9. Знайти довірчий інтервал для параметрів гіпотетичного закону розподілу, взяв рівень надійності $\gamma = 0.95$.

Оскільки для параметра a довірчий інтервал побудувати простіше, я почну саме з нього.
1. Параметр a
Згадаємо, що оцінка $a^* = \min\limits_{1 \leq k \leq n} \xi_k$, ми знаємо її розподіл $a^* \sim Exp\left(\frac{n}{\lambda}\ ,\ a\right)$, а отже $a^* \geq a$. Шукатимемо однобічний довірчий інтервал з умови $P\{a \in (a^* - \varepsilon\ ;\ a^*]\} = \gamma$.

$$a \in (a^* - \varepsilon\ ;\ a^*] \iff a^* - \varepsilon\ < a\ \leq a^* \iff a\ \leq\ a^*\ <\ a + \varepsilon$$

$$F_{a^*}(x) = \begin{cases} 1 - e^{-\frac{n(x-a)}{\lambda}}, & x > a \\ \\ 0, & x \leq a \end{cases}$$

Маємо:

$$P\{a\ <\ a^*\ \leq\ a + \varepsilon\} = F_{a^*}(a + \varepsilon) - F_{a^*}(a) = 1 - e^{-\frac{n(a+\varepsilon-a)}{\lambda}} - 1 + e^{\frac{n(a-a)}{\lambda}} =$$

$$= 1 - e^{-\frac{n\varepsilon}{\lambda}} = \gamma$$

За значення параметра λ приймемо значення його точкової оцінки $\lambda^{**} = 3.1999$, $n = 100$, $\gamma = 0.95$. Розв'яжемо отримане рівняння відносно ε:

$$1 - e^{-\frac{100\varepsilon}{3.1999}} = 0.95 \ \Rightarrow\ e^{-\frac{100\varepsilon}{3.1999}} = 0.05 \ \Rightarrow\ -\frac{100\varepsilon}{3.1999} = \ln 0.05 \ \Rightarrow$$

$$\Rightarrow\ \varepsilon = -\frac{3.1999}{\ln 0.05 \cdot 100} = 0.01068$$

Тепер, маючи значення ε отримали довірчий інтервал для параметра a:

$$a\ \in\ (2.55932\ ;\ 2.57]$$

2. Параметр λ
Оскільки закон розподілу оцінки параметра λ нам невідомий, а дослідити λ^{**} на асимптотичну нормальність досить важко спробуємо піти іншим шляхом. Згідно до статті 3:

$$\sqrt{I\,(\theta)}\sqrt{n}\,(\theta^* - \theta) \xrightarrow{F}\ \eta\ \sim\ N(0,1)$$

Тут у якості оцінки параметра беремо асимптотично незміщену оцінку (тобто, безпосередньо ту, що отримали методом максимальної правдоподібності). Отже, для нашого випадку маємо:

$$\frac{n}{\lambda}\,(\lambda^* - \lambda) \xrightarrow{F}\ \eta\ \sim\ N(0,1)$$

Тепер можемо побудувати наближений довірчий інтервал. Маємо рівність:

$$P\{\frac{n}{\lambda}|\lambda^* - \lambda| < t_\gamma\} = \gamma \;\;\Rightarrow\;\; P\{\frac{n}{\lambda}|\lambda^* - \lambda| < t_\gamma\} = 2\Phi\big(t_\gamma\big) = \gamma$$

Користуючись таблицею значень функції Лапласа знайдемо t_γ :

$$2\Phi\big(t_\gamma\big) = 0.95 \;\;\Rightarrow\;\; \Phi\big(t_\gamma\big) = 0.475 \;\;\Rightarrow\;\; t_\gamma \approx 1.96$$

Тепер розв'яжемо нерівність відносно λ:

$$\frac{n}{\lambda}|\lambda^* - \lambda| < t_\gamma \;\;\Rightarrow\;\; |\lambda^* - \lambda| < \frac{\lambda t_\gamma}{n} \;\;\Rightarrow\;\; (\lambda^* - \lambda)^2 < \frac{\lambda^2 t_\gamma^2}{n^2} \;\;\Rightarrow$$

$$\Rightarrow\;\; \left(\lambda^*\right)^2 - 2\lambda\lambda^* + \lambda^2 < \frac{\lambda^2 t_\gamma^2}{n^2} \;\;\Rightarrow\;\; \left(\lambda^*\right)^2 - 2\lambda\lambda^* + \lambda^2 - \frac{\lambda^2 t_\gamma^2}{n^2} < 0 \;\;\Rightarrow$$

$$\Rightarrow\;\; \left(\lambda^*\right)^2 - 2\lambda\lambda^* + \lambda^2\left(1 - \frac{t_\gamma^2}{n^2}\right) < 0$$

Знайдемо корені квадратного рівняння, позначивши $1 - \frac{t_\gamma^2}{n^2} = b$:

$$b\lambda^2 - 2\lambda\lambda^* + \left(\lambda^*\right)^2 = 0 \;\;\Rightarrow\;\; \left[\; D = 4(\lambda^*)^2 - 4b(\lambda^*)^2 = 4(\lambda^*)^2\,(1-b) \;\right] \;\;\Rightarrow\;\; \lambda_{1,2} = \frac{2\lambda^* \pm 2\lambda^*\sqrt{1-b}}{2b}$$

Отже, повернувшись до заміни отримали, що довірчий інтервал має вигляд:

$$(\lambda_1, \lambda_2) = \left(\frac{2\lambda^* - 2\lambda^*\sqrt{1 - 1 + \frac{t_\gamma^2}{n^2}}}{2\left(1 - \frac{t_\gamma^2}{n^2}\right)} \;;\; \frac{2\lambda^* + 2\lambda^*\sqrt{1 - 1 + \frac{t_\gamma^2}{n^2}}}{2\left(1 - \frac{t_\gamma^2}{n^2}\right)}\right) = \left(\frac{2\lambda^*\left(1 - \frac{t_\gamma}{n}\right)}{2\left(1 - \frac{t_\gamma^2}{n^2}\right)} \;;\; \frac{2\lambda^*\left(1 + \frac{t_\gamma}{n}\right)}{2\left(1 - \frac{t_\gamma^2}{n^2}\right)}\right) =$$

$$= \left(\frac{\lambda^*}{1 + \frac{t_\gamma}{n}} \;;\; \frac{\lambda^*}{1 - \frac{t_\gamma}{n}}\right)$$

Значення λ^* обчислимо як:

$$\left(\lambda^*\right)_{\text{зн}} = \frac{n-1}{n}\left(\lambda^{**}\right)_{\text{зн}} = \frac{99 \cdot 3.1999}{100} \approx 3.1679$$

Отже, маємо наближений довірчий інтервал:

$$\lambda \;\in\; (3.1070 \;;\; 3.2312)$$

10. Висновки.

Отже, під час виконання даної розрахункової роботи було досліджено задану конкретну реалізацію вибірки зі ста елементів. В першу чергу виконали первинну обробку даних: вибірку було відсортовано, далі, оскільки майже усі значення були унікальними, було побудовано інтервальний варіаційний ряд та висунуто припущення, що генеральна сукупність, з якої отримали конкретну реалізацію вибірки, має неперервний розподіл. Далі, побудували графічне зображення варіаційного ряду - гістограму та її обвідну. Оскільки гістограма, в деякому сенсі, є аналогом щільності розподілу, її вигляд та вигляд "зглаженої гістограми" згодом допоміг висунути гіпотезу про закон розподілу ГС.
Далі, було побудовано емпіричну функцію розподілу для конкретної реалізації вибірки. Оскільки я підозрювала, що ГС в результаті матиме неперервний розподіл, окрім графіка ЕФР було побудовано і кумулятивну криву, яка була схожа до графіка функції розподілу експоненційного закону розподілу зі зсувом, проте даних для висування гіпотези усе ще було недостатньо.
Наступним кроком було обчислено значення вибіркової медіани моди та асиметрії. Результати отримані під час обчислень підтвердили мої підозри стосовно закону розподілу. Так, медіана лежала ближче до лівого краю вибірки, модальний інтервал був чітко виражений, а саме значення моди було досить близьке до мінімуму вибірки, що є характерним для експоненційного закону. Асиметрія, хоча й не дорівнювала чітко двом, була значно більша за нуль і доволі близька до двох. Також знайшли незміщені оцінки математичного сподівання та дисперсії.
Отже, наступним кроком було висунуто гіпотезу H_0 про можливий закон розподілу ГС з якої узято дану конкретну реалізацію вибірки,
$H_0 : \quad \xi \sim \quad Exp\left(\frac{1}{\lambda}, a\right)$. Також я висунула і альтернативну гіпотезу (на випадок, якщо не підтвердиться основна) - ГС розподілена за іншим законом.
Для виконання подільшої роботи, в тому числі і для перевірки основної гіпотези, потрібно було оцінити параметри закону розподілу. Я навела оцінки методом моментів та методом максимальної правдоподібності, однак оцінки отримані другим методом були значно простішими у дослідженні, тому в подальшому я використала саме їх. Слід зазначити, що оцінку для параметра a було знайдено з урахуванням властивостей самого параметра, оскільки вираз $\frac{\partial \ln \mathcal{L}_{Exp}(\overrightarrow{x}, \lambda, a)}{\partial a} = \frac{n}{\lambda}$ не набуває нульових значень, а отже і критичних значень параметра a теж немає.
У наступному пункті відбулася перевірка основної гіпотези H_0. Для цього прийняли значення точкових оцінок знайдених раніше за наближене значення парметрів. Після перевірки я зробила висновок, що на рівні значущості $\alpha = 0.05$ дані не суперечать висунутій гіпотезі. Зрештою, були побудовані довірчі інтервали для парметрів гіпотетичного закону розподілу. Проте, межі цих інтервалів не можна вважати точними, оскільки замість невідомого значення параметра λ було використане значення його точкової оцінки (під час пошуку довірчого інтервалу для параметра a), а пошук довірчого інтервалу для параметра λ проводився для його асимптотично незміщенної оцінки.

Отже, на рівні значущості $\alpha = 0.05$ дані не суперечать висунутій гіпотезі H_0. Однак, стверджувати, що ГС з якої узято конкретну реалізацію вибірки, розподілена саме за експоненційним законом зі зсувом не можна. Це пов'язано з тим, що при виконанні даної роботи я розглядала лише канонічні закони розподілу для неперервних випадкових величин, хоча ГС може бути розподілена і за іншим законом, характеристики якого співпадають з отриманими під час первинної обробки даних.

СПИСОК ЛІТЕРАТУРИ

1. МГТУ им. Баумана - "Математическая статистика", Москва 2001
2. Journal of Probability and Statistics, Volume 2012, Article ID734575 - "Interval Estimations of the Two-Parameter Exponential Distribution" (https://www.hindawi.com/journals/jps/2012/734575/)
3. https://web.stanford.edu/class/stats200/Lecture14.pdf
4. Гаральд Крамер - "Математические методы статистики", Москва 1975
5. Лагутин М.Б. - "Наглядная математическая статистика", Москва 2019
6. Електронний конспект лекцій.