

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
“ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ”
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ”**

**РОЗРАХУНКОВА ГРАФІЧНА РОБОТА
З КУРСУ "МАТЕМАТИЧНА СТАТИСТИКА"
Варіант: 104**

Виконав:
Вітковський Д.О.

Прийняв:
Ільєнко А.Б.

Київ - 2021

Зміст

1. Первинний аналіз вибірки.	5
2. Знаходження описових статистик.	10
3. Гіпотеза про розподіл	12
4. Оцінка параметрів.	14
5. Дослідження оцінки.	16
6. Інтервальні оцінки параметрів.	19
7. Перевірка гіпотези про розподіл.	22
8. Висновки.	23

Завдання.

- 1) Провести первинний аналіз вибірки. Це включає статистичний ряд, емпіричну функцію розподілу, її графік, полігон частот, гістограму, box-and-whiskers plot.
- 2) Знайти вибіркове середнє, вибіркиму дисперсію, виправлену вибіркиму дисперсію, вибіркиму медіану, вибіркиму моду, вибіркові коефіцієнти асиметрії та ексцесу.
- 3) Обґрунтувати та висунути (нову) гіпотезу про розподіл генеральної сукупності.
- 4) Методом моментів та методом максимальної вірогідності знайти оцінки параметрів розподілу.
- 5) Для кожного з параметрів кращу з цих двох оцінок перевірити на (асимптотичну незміщеність), консистентність та ефективність.
- 6) Побудувати довірчі інтервали надійністю 0.95 для параметрів розподілу.
- 7) Перевірити висунуту гіпотезу про розподіл генеральної сукупності за допомогою критерію χ^2 . Якщо гіпотеза суперечить вибірким даним, перейти до п. 3.
- 8) Зробити висновки.

Вихідна реалізація вибірки:

-1.66	-0.10	2.84	5.04	1.60	-2.80	-0.72	5.47	-4.20	0.67
-2.49	-2.40	5.99	-3.26	2.16	0.51	-3.49	5.10	-1.24	2.58
-1.06	-0.84	4.09	6.54	-2.96	6.77	7.29	-2.87	2.29	1.37
2.51	1.67	-4.48	-1.46	0.84	-2.01	-1.07	0.01	-4.33	3.41
0.49	-3.41	9.65	-3.84	6.15	8.17	-1.55	-3.60	-2.73	18.49
1.38	6.09	-2.15	9.68	-0.47	7.67	-1.47	3.30	4.58	0.43
2.19	5.19	1.95	-4.44	-0.22	-4.49	-3.06	1.09	-1.03	-1.18
3.52	2.15	-3.48	3.64	-3.21	-0.82	-3.29	-2.09	2.81	-0.92
0.24	1.39	-4.22	1.20	-2.68	-1.93	1.49	-3.64	1.58	1.59
0.66	-2.04	3.41	2.69	3.73	-2.11	4.02	-0.66	1.25	-3.07

Відсортована реалізація вибірки:

-4.49	-4.48	-4.44	-4.33	-4.22	-4.20	-3.84	-3.64	-3.60	-3.49
-3.48	-3.41	-3.29	-3.26	-3.21	-3.07	-3.06	-2.96	-2.87	-2.80
-2.73	-2.68	-2.49	-2.40	-2.15	-2.11	-2.09	-2.04	-2.01	-1.93
-1.66	-1.55	-1.47	-1.46	-1.24	-1.18	-1.07	-1.06	-1.03	-0.92
-0.84	-0.82	-0.72	-0.66	-0.47	-0.22	-0.1	0.01	0.24	0.43
0.49	0.51	0.66	0.67	0.84	1.09	1.20	1.25	1.37	1.38
1.39	1.49	1.58	1.59	1.60	1.67	1.95	2.15	2.16	2.19
2.29	2.51	2.58	2.69	2.81	2.84	3.30	3.41	3.41	3.52
3.64	3.73	4.02	4.09	4.58	5.04	5.10	5.19	5.47	5.99
6.09	6.15	6.54	6.77	7.29	7.67	8.17	9.65	9.68	18.49

1. Первинний аналіз вибірки.

Оскільки з даних видно, що розподіл неперервний, то для побудови статистичного ряду дані з вибірки треба розбити на рівновеликі інтервали. Емпіричною формулою кількості таких інтервалів є формула Стерджеса:

$$m = 1 + \log_2 n = 1 + \log_2 100 = 7.644 \approx 8. \quad (1)$$

Але мені така кількість видається замалою, тому використаю значення

$$m = 12. \quad (2)$$

Отже

$$x_{\min} = \min\{x\} = -4.49;$$

$$x_{\max} = \max\{x\} = 18.49.$$

Розмах вибірки

$$R = x_{\max} - x_{\min} = 22.98.$$

Таким чином довжина кроку дорівнює 1.915.

Інтервал	$[-4.49, -2.575)$	$[-2.575, -0.66)$	$[-0.66, 1.255)$	$[1.255, 3.17)$
Середина інтервалу	-3.525	-1.6175	0.2975	4.425
Частота	22	21	15	18
Кумулятивна частота	22	43	58	76
Відносна частота	0.22	0.21	0.15	0.18
Кумулятивна відносна частота	0.22	0.43	0.58	0.76

Таблиця 1. Статистичний ряд.

Інтервал	[3.17, 5.085)	[5.085, 7)	[7, 8.915)	[8.915, 10.83)
Середина інтервалу	4.1275	6.0425	7.9575	9.8725
Частота	10	8	3	2
Кумулятивна частота	86	94	97	99
Відносна частота	0.1	0.08	0.03	0.02
Кумулятивна відносна частота	0.86	0.91	0.97	0.99

Інтервал	[10.83, 12.745)	[12.745, 14.66)	[14.66, 16.575)	[16.575, 18.49]
Середина інтервалу	11.7875	13.7025	15.6175	17.5325
Частота	0	0	0	1
Кумулятивна частота	99	99	99	100
Відносна частота	0	0	0	0.01
Кумулятивна відносна частота	0.99	0.99	0.99	1

Таблиця 2. Статистичний ряд (продовження).

Емпірична функція розподілу (інтервальна):

$$F_n^*(x) = \begin{cases} 0, & x < x_0^*; \\ \omega_k^{\text{кумуля}}, & x_k^* \leq x < x_{k+1}^*. \end{cases} \quad (3)$$

, де x_k^* – початок k-того інтервалу, кінець (k-1)-го, а $\omega_k^{\text{кумуля}}$ – кумулятивна відносна частота для k-того інтервалу.

Відповідна функція для даного інтервального розбиття:

$$F_n^*(x) = \begin{cases} 0, & x < -4.49; \\ 0.22, & x \in [-4.49, -2.575); \\ 0.43, & x \in [-2.575, -0.66); \\ 0.58, & x \in [-0.66, 1.255); \\ 0.76, & x \in [1.255, 3.17); \\ 0.86, & x \in [3.17, 5.085); \\ 0.91, & x \in [5.085, 7); \\ 0.97, & x \in [7, 8.915); \\ 0.99, & x \in [8.915, 16.575); \\ 1, & x \geq 16.575. \end{cases} \quad (4)$$

Її графік:

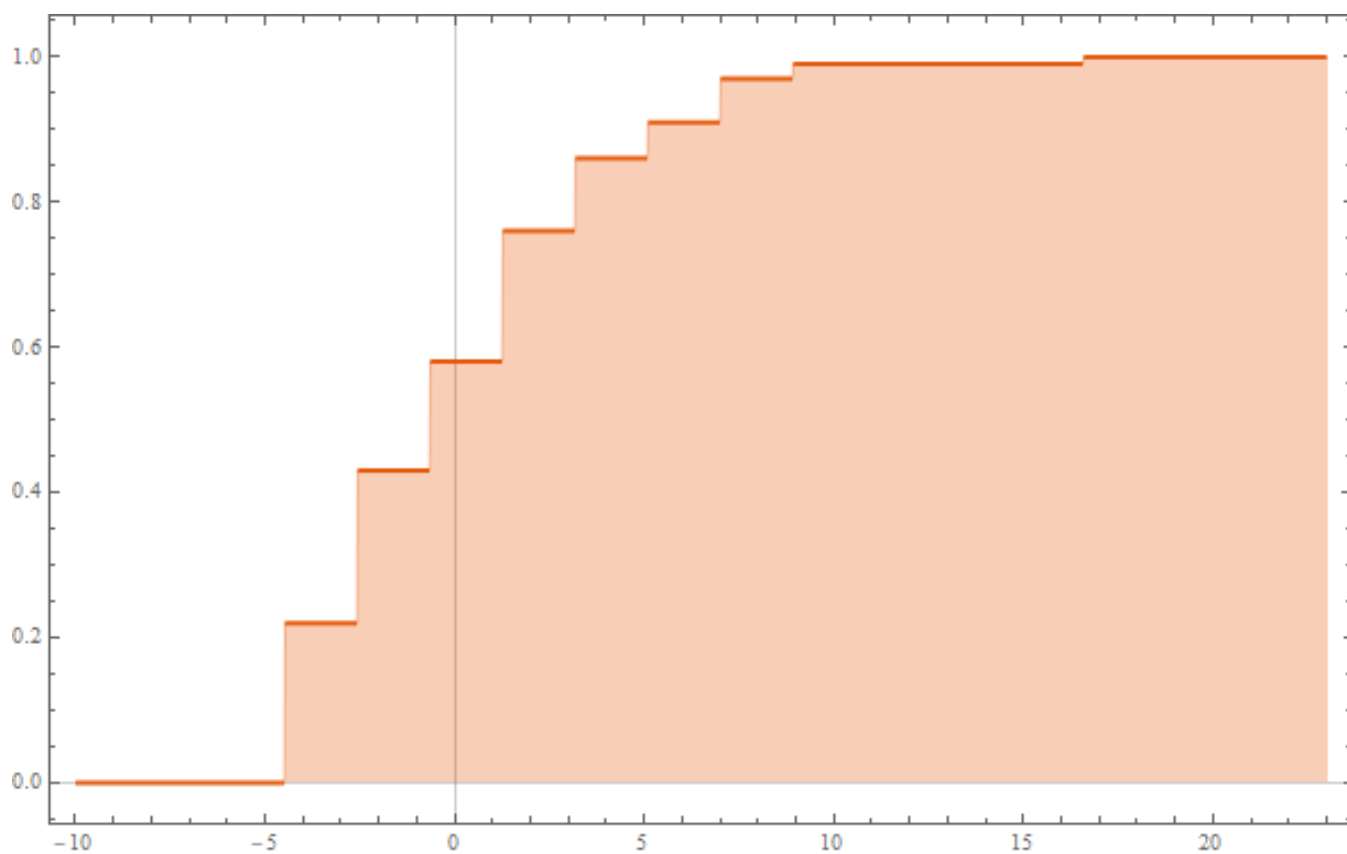


Рис. 1. Інтервальна емпірична функція розподілу.

Гістограма за заданими інтервалами:

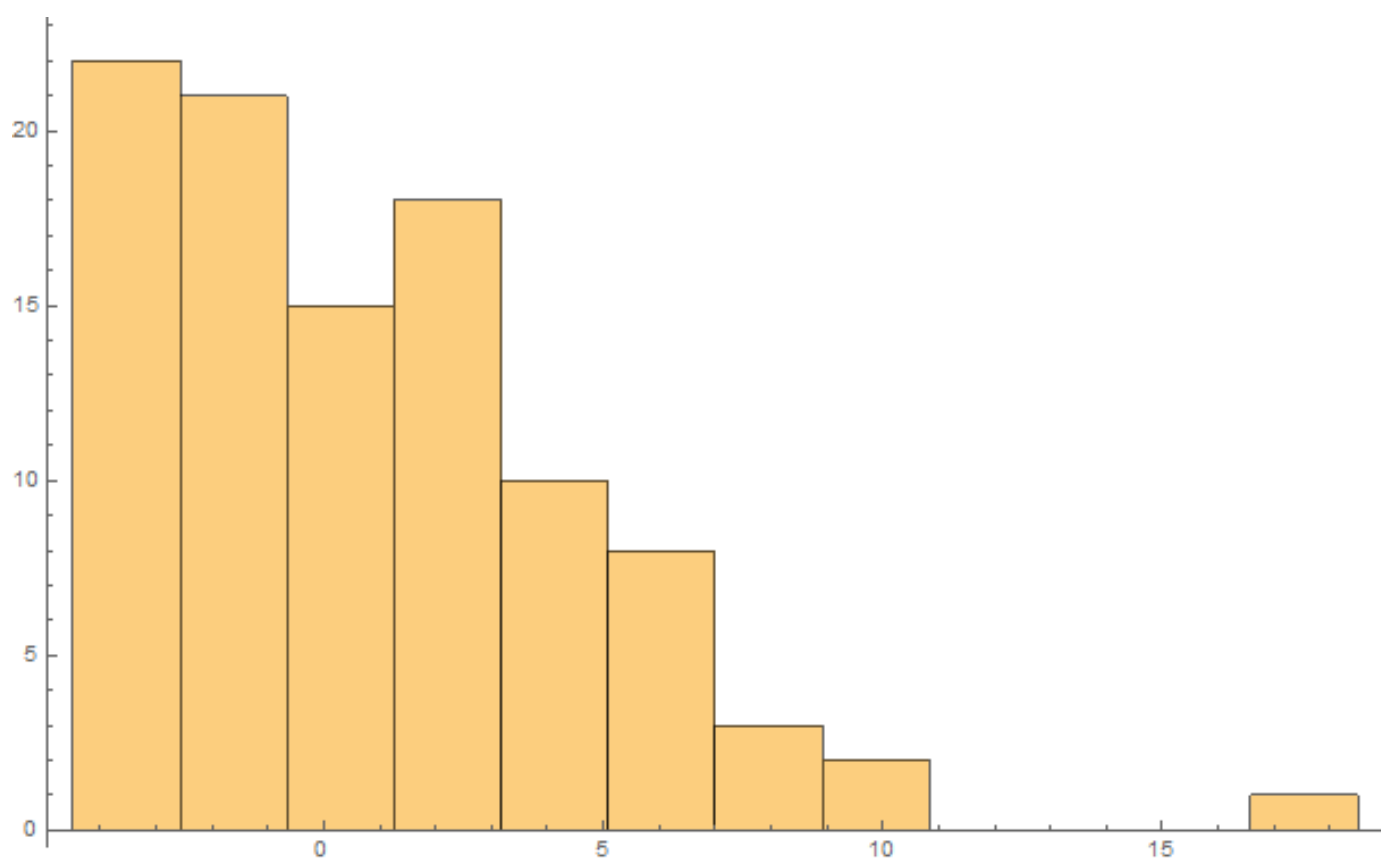


Рис. 2. Гістограма розподілу.

Для побудови графіку "ящик з вусами" використано значення, обчислені у наступному розділі, а саме: вибіркове середнє, медіану, перший та третій квартилі і мінімум та максимум як значення кінців вусів.

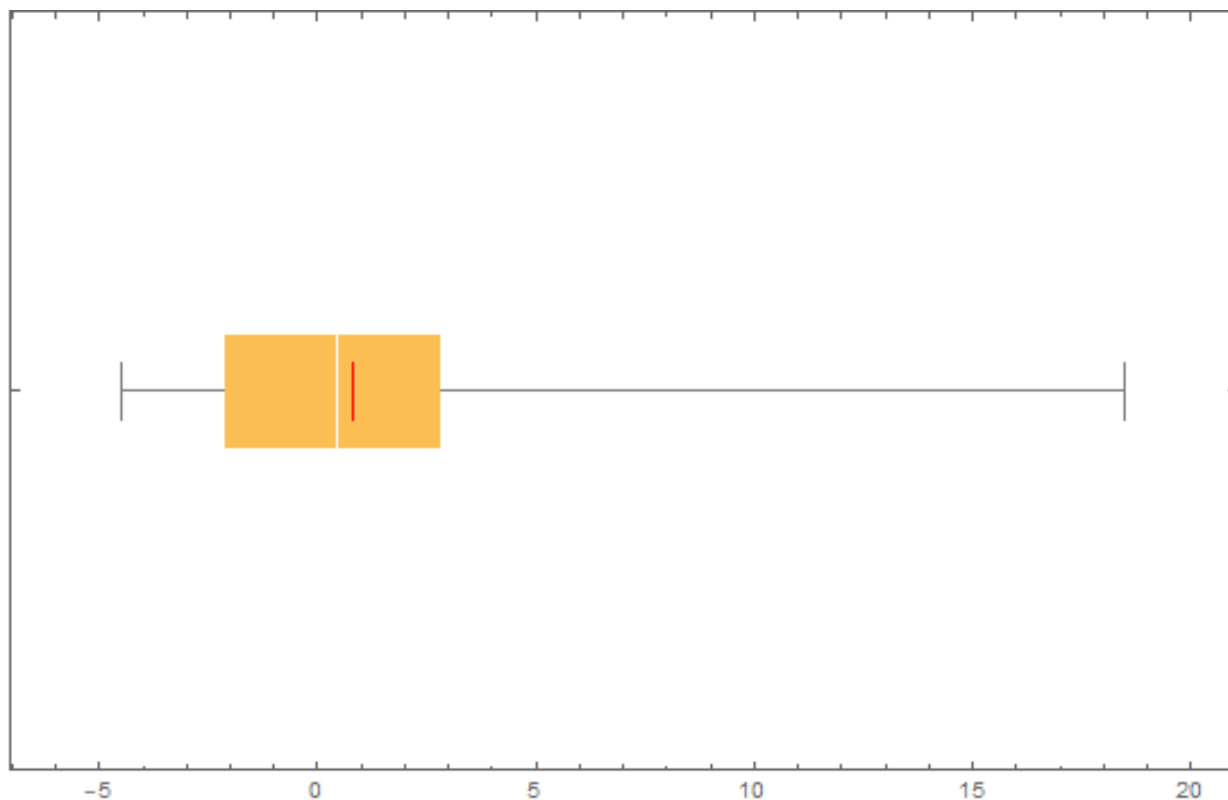


Рис. 3. Ящик з вусами.

2. Знаходження описових статистик.

Вибіркове середнє:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{100} \sum_{i=1}^{100} x_i = 0.7938 \quad (5)$$

Вибіркова дисперсія:

$$s^2 = \mathbb{D}_{\xi}^{**} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{100} \sum_{i=1}^{100} (x_i - 0.7938)^2 = 15.0129 \quad (6)$$

Виправлена вибірка дисперсія:

$$s_0^2 = \mathbb{D}_{\xi}^{***} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{99} \sum_{i=1}^{100} (x_i - 0.7938)^2 = 15.1645 \quad (7)$$

Вибіркова медіана:

Оскільки маємо відсортований набір усіх x , можна доволі легко обчислити медіану за наступною формулою:

$$\mathcal{M}_e^* = \begin{cases} x_{(k+1)}, & n = 2k + 1 \\ \frac{x_{(k)} + x_{(k+1)}}{2}, & n = 2k \end{cases}$$

, де $k \in \mathbb{N}_0$

Тоді у нашому випадку медіана обчислюється наступним чином:

$$\mathcal{M}_e^* = \frac{x_{(k)} + x_{(k+1)}}{2} = \frac{0.43 + 0.49}{2} = 0.46 \quad (8)$$

Вибіркова мода:

$$mo = 1 - \text{номер модального класу} \quad (9)$$

$$\mathcal{M}_o^* = y_{mo-1} + (y_{mo} - y_{mo-1}) \cdot \frac{n_{mo} - n_{mo-1}}{(n_{mo} - n_{mo-1}) + (n_{mo} - n_{mo+1})} \quad (10)$$

, де y_{mo-1} , y_{mo} – відповідно нижня та верхня межі модального класу; n_{mo} , n_{mo-1} , n_{mo+1} – частоти відповідно модального, передмодального та післямодального інтервалів.

$$\boxed{\equiv} - 4.49 + (-2.575 - (-4.49)) \cdot \frac{22 - 0}{(22 - 0) + (22 - 21)} = -2.65826$$

Вибіркові коефіцієнти асиметрії та екцесу:

Для знаходження цих коефіцієнтів потрібно обчислювати вибіркові центральні моменти за наступною формулою:

$$\overline{\mu}_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k \quad (11)$$

Вибірковий коефіцієнт асиметрії:

$$As = \frac{\overline{\mu}_3}{s_0^3} = \frac{\frac{1}{100} \sum_{i=1}^{100} (x_i - \bar{x})^3}{\left(\mathbb{D}_\xi^{***}\right)^{\frac{3}{2}}} = 1.22459 \quad (12)$$

Вибірковий коефіцієнт екцесу:

$$Ek = \frac{\overline{\mu}_4}{s_0^4} - 3 = \frac{\frac{1}{100} \sum_{i=1}^{100} (x_i - \bar{x})^4}{\left(\mathbb{D}_\xi^{***}\right)^2} - 3 = 2.822 \quad (13)$$

Перший кuartиль:

$$\begin{aligned} \alpha = 0.25 &\Rightarrow K = n \cdot \alpha = 100 * 0.25 = 25 \in \mathbb{Z} \Rightarrow \\ \Rightarrow Q_1 = q_\alpha &= \frac{x_{(K)} + x_{(K+1)}}{2} = \frac{-2.15 - 2.11}{2} = -2.13 \end{aligned} \quad (14)$$

Третій кuartиль:

$$\begin{aligned} \alpha = 0.75 &\Rightarrow K = n \cdot \alpha = 100 * 0.75 = 75 \in \mathbb{Z} \Rightarrow \\ \Rightarrow Q_3 = q_\alpha &= \frac{x_{(K)} + x_{(K+1)}}{2} = \frac{2.81 - 2.84}{2} = 2.825 \end{aligned} \quad (15)$$

3. Гіпотеза про розподіл

Дані вибірки майже не мають повторень та не є цілими числами, тому розподіл є неперервним. Вибірка найбільше схожа на зміщену експоненційно розподілену, оскільки має найбільшу частоту у лівій своїй частині яка швидко зменшується при русі у додатному напрямку. Її гістограма та емпірична функція розподілу схожі на відповідні функції для зміщеного на 4.49 у від'ємному напрямку експоненціального розподілу.

Таким чином висунемо гіпотезу що ГС розподілена за експоненційним законом зі зміщенням a та відповідним параметром $\frac{1}{\lambda}$, де $\lambda = \mathbb{E}_\xi - a$:

$$H_0 = \left\{ \xi \sim \text{Exp}(\lambda, a) = \text{Exp}\left(\frac{1}{\lambda}\right) + a \right\}$$

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

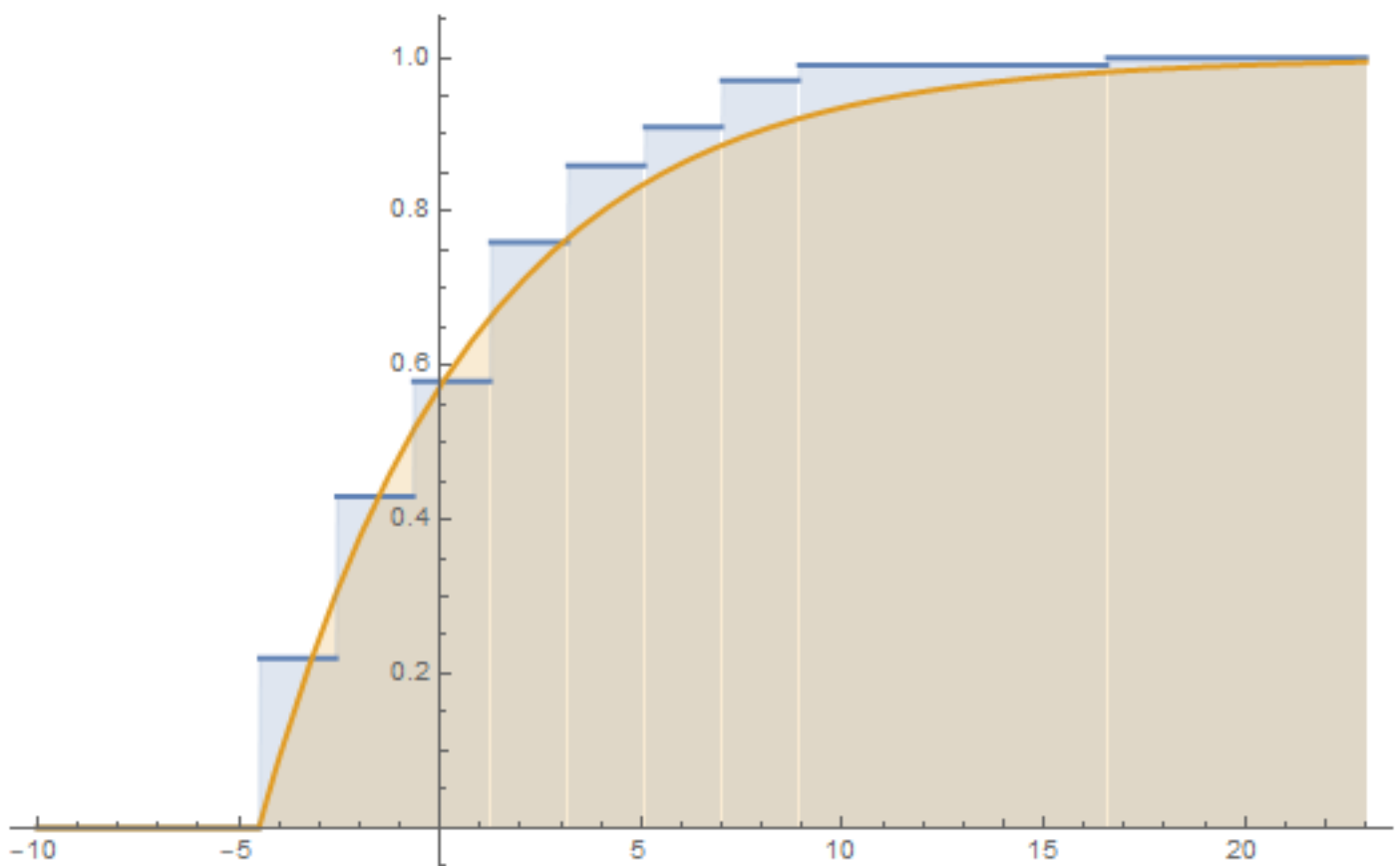


Рис. 4. Емпірична функція розподілу у порівнянні з відповідною функцією розподілу.

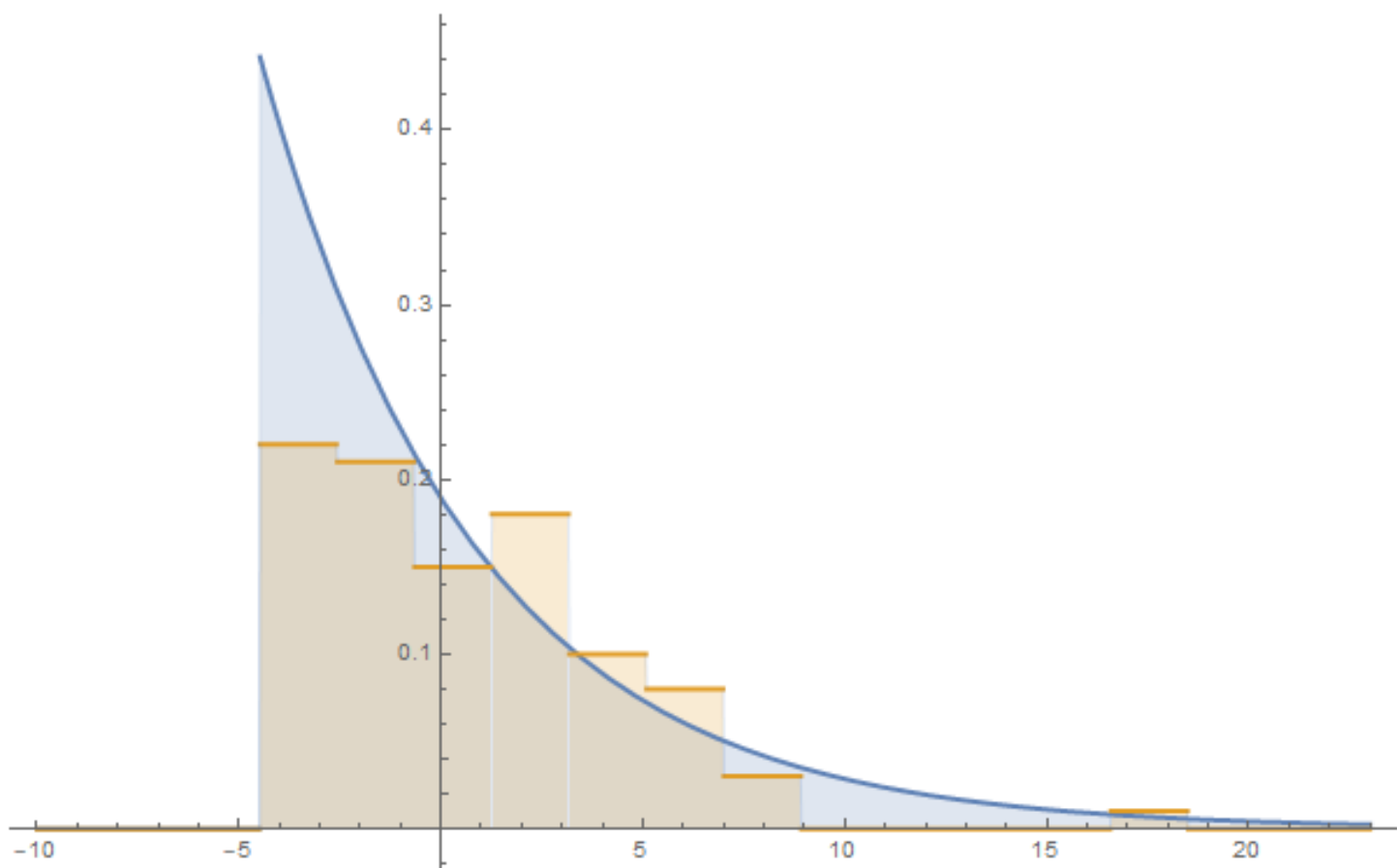


Рис. 5. Гістограма у порівнянні із відповідною функцією щільності.

Тоді альтернативною гіпотезою буде $H_1 = \{\xi \approx \text{Exp}(\lambda, a)\}$.

4. Оцінка параметрів.

Для зміщеного експоненційного розподілу маємо параметри λ та a . Треба їх оцінити.

Метод моментів.

Для знаходження оцінки методом моментів, треба розв'язати систему рівнянь, що включає s моментів розподілу, де s – кількість параметрів розподілу, а також їх емпіричні еквіваленти.

$$\begin{aligned}\mathbb{E}\xi &= \int_{-\infty}^{+\infty} x f_{\xi}(x) dx = \int_a^{+\infty} x \cdot \frac{1}{\lambda} e^{-\frac{x-a}{\lambda}} dx = \frac{1}{\lambda} \int_a^{+\infty} x e^{-\frac{x-a}{\lambda}} dx = \\ &= \left| \begin{array}{l} u = x; \quad dv = e^{-\frac{x-a}{\lambda}} dx \\ du = dx; \quad v = -\lambda e^{-\frac{x-a}{\lambda}} \end{array} \right| = \frac{1}{\lambda} \left(-\lambda x e^{-\frac{x-a}{\lambda}} \Big|_a^{+\infty} + \lambda \int_a^{+\infty} e^{-\frac{x-a}{\lambda}} dx \right) = \\ &= \frac{1}{\lambda} \left(0 - (-\lambda a) - \lambda^2 e^{-\frac{x-a}{\lambda}} \Big|_a^{+\infty} \right) = \frac{1}{\lambda} (\lambda a - \lambda^2 (0 - 1)) = \boxed{a + \lambda}; \quad (16)\end{aligned}$$

$$\begin{aligned}\mathbb{E}\xi^2 &= \int_{-\infty}^{+\infty} x^2 \cdot f_{\xi}(x) dx = \frac{1}{\lambda} \int_a^{+\infty} x^2 e^{-\frac{x-a}{\lambda}} dx = \left| \begin{array}{l} u = x^2; \quad dv = e^{-\frac{x-a}{\lambda}} dx \\ du = 2x dx; \quad v = -\lambda e^{-\frac{x-a}{\lambda}} \end{array} \right| = \\ &= \frac{1}{\lambda} \left(-\lambda x^2 e^{-\frac{x-a}{\lambda}} \Big|_a^{+\infty} + 2\lambda \int_a^{+\infty} x e^{-\frac{x-a}{\lambda}} dx \right) = \frac{1}{\lambda} (\lambda a^2 + 2\lambda (\lambda a + \lambda^2)) = \\ &= a^2 + 2\lambda a + 2\lambda^2 = \boxed{(a + \lambda)^2 + \lambda^2} \quad (17)\end{aligned}$$

Отримуємо наступну систему рівнянь:

$$\begin{cases} a^* + \lambda^* = \bar{\xi} \\ (a^* + \lambda^*)^2 + (\lambda^*)^2 = \bar{\xi}^2 \end{cases}$$
$$\begin{cases} \lambda^* = \sqrt{\bar{\xi}^2 - \bar{\xi}^2} \approx 3.0174 \\ a^* = \bar{\xi} - \sqrt{\bar{\xi}^2 - \bar{\xi}^2} \approx -2.2236 \end{cases} \quad (18)$$

Метод максимальної вірогідності.

Для знаходження оцінки методом максимальної вірогідності треба максимізувати функцію вірогідності за параметрами λ та a . Можемо спробувати знайти її глобальний максимум.

$$\begin{aligned}\mathcal{L}(x_1, \dots, x_n, \lambda, a) &= f_{\xi_1, \dots, \xi_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{\xi_i}(x_i) = \prod_{i=1}^n f_{\xi}(x_i) = \\ &= |x \geq a| = \prod_{i=1}^n \frac{1}{\lambda} e^{-\frac{x_i - a}{\lambda}} = \frac{1}{\lambda^n} e^{-\frac{1}{\lambda}(x_1 + \dots + x_n - na)}\end{aligned}$$

Для легшого обчислення похідних варто використати функцію логарифма, оскільки вона монотонно зростаюча, а логарифм добутку родівнює сумі логарифмів множників, що спрощує диференціювання.

$$\ln \mathcal{L} = -n \ln \lambda - \frac{x_1 + \dots + x_n - na}{\lambda}$$

Знайдемо часткові похідні першого та другого порядків за параметрами:

$$\frac{\partial \ln \mathcal{L}}{\partial \lambda} = -\frac{n}{\lambda} + \frac{x_1 + \dots + x_n - na}{\lambda^2} = \frac{x_1 + \dots + x_n - na - n\lambda}{\lambda^2};$$

$$\frac{\partial \ln \mathcal{L}}{\partial a} = \frac{n}{\lambda};$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda^2} = \frac{n}{\lambda^2} - 2 \frac{x_1 + \dots + x_n - na}{\lambda^3};$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial a^2} = 0;$$

$$\frac{\partial^2 \ln \mathcal{L}}{\partial \lambda \partial a} = -\frac{n}{\lambda^2};$$

$$\det \begin{pmatrix} \frac{\partial^2 \ln \mathcal{L}}{\partial \lambda^2} & \frac{\partial^2 \ln \mathcal{L}}{\partial \lambda \partial a} \\ \frac{\partial^2 \ln \mathcal{L}}{\partial \lambda \partial a} & \frac{\partial^2 \ln \mathcal{L}}{\partial a^2} \end{pmatrix} = 0 - \left(-\frac{n}{\lambda^2}\right)^2 < 0$$

Отже функція не має екстремумів, тому принаймні спробуємо максимізувати її значення.

$$\begin{cases} \frac{\partial \ln \mathcal{L}}{\partial \lambda^*} = 0; \\ \frac{\partial \ln \mathcal{L}}{\partial a^*} = 0. \end{cases} \Rightarrow \begin{cases} a^* + \lambda^* = \bar{x}; \\ \lambda^* \neq 0; \\ \lambda^* \rightarrow \infty. \end{cases} \Rightarrow \begin{cases} a^* = \min\{x_i\} = -4.49; \\ \lambda^* = \bar{x} - a^* = 5.2838. \end{cases} \quad (19)$$

5. Дослідження оцінки.

Кращою з двох отриманих оцінок є оцінка методом максимальної вірогідності. Проведемо її дослідження.

Незміщеність.

$$\mathbb{E}a^* = \mathbb{E}\min\{\xi_i\} \equiv$$

Щільність розподілу мінімуму:

$$\begin{aligned} f_\xi(x) &= \frac{1}{\lambda} e^{-\frac{x-a}{\lambda}} \cdot \mathbb{I}\{x \geq a\}; \quad F_\xi(x) = (1 - e^{-\frac{x-a}{\lambda}}) \cdot \mathbb{I}\{x \geq a\} \\ f_{\min\{\xi\}}(x) &= n(1 - F_\xi(x))^{n-1} f_\xi(x) = \frac{n}{\lambda} e^{-n\frac{x-a}{\lambda}} \cdot \mathbb{I}\{x \geq a\} \\ &\equiv \int_a^{+\infty} x \cdot \frac{n}{\lambda} e^{-n\frac{x-a}{\lambda}} dx = \left| \begin{array}{l} u = x; \quad dv = e^{-n\frac{x-a}{\lambda}} dx \\ du = dx; \quad v = -\frac{\lambda}{n} e^{-n\frac{x-a}{\lambda}} \end{array} \right| = \\ &\frac{n}{\lambda} \left(-\frac{\lambda}{n} x e^{-n\frac{x-a}{\lambda}} \Big|_a^{+\infty} + \frac{\lambda}{n} \int_a^{+\infty} e^{-n\frac{x-a}{\lambda}} dx \right) = a - \frac{\lambda}{n} e^{-n\frac{(x-a)}{\lambda}} \Big|_a^{+\infty} = \frac{na + \lambda}{n} \neq a \Rightarrow \\ &\Rightarrow a^* - \text{не є незміщеною.} \end{aligned} \quad (20)$$

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}a^* &= \lim_{n \rightarrow \infty} \frac{na + \lambda}{n} = a \Rightarrow \\ &\Rightarrow a^* - \text{є асимптотично незміщеною.} \end{aligned} \quad (21)$$

$$\begin{aligned} \mathbb{E}\lambda^* &= \mathbb{E}(\bar{\xi} - a^*) = \mathbb{E}\frac{\xi_1 + \dots + \xi_n}{n} - \frac{na + \lambda}{n} = \frac{1}{n} n \mathbb{E}\xi_i - \frac{na + \lambda}{n} = \\ &a + \lambda - \frac{na + \lambda}{n} = \frac{n\lambda - \lambda}{n} \neq \lambda \Rightarrow \\ &\Rightarrow \lambda^* - \text{не є незміщеною.} \end{aligned} \quad (22)$$

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{E}\lambda^* &= \lim_{n \rightarrow \infty} \frac{n\lambda - \lambda}{n} = \lambda \Rightarrow \\ &\Rightarrow \lambda^* - \text{є асимптотично незміщеною.} \end{aligned} \quad (23)$$

Консистентність.

$$\mathbb{P} \lim_{n \rightarrow \infty} a^* \stackrel{?}{=} a$$

$$\mathbb{P} \lim_{n \rightarrow \infty} \lambda^* \stackrel{?}{=} \lambda$$

Перевіримо за лемою про консистентність. Для цього потрібно обчислити дисперсії оцінок. Пізніше вона усе одно знадобиться для обчислення інтервальних оцінок.

$$\mathbb{D}a^* = \mathbb{E}(a^*)^2 - (\mathbb{E}a^*)^2 ;$$

$$\begin{aligned} \mathbb{E}(a^*)^2 &= \frac{n}{\lambda} \int_a^{+\infty} x^2 \cdot e^{-n \frac{x-a}{\lambda}} dx = \left| \begin{array}{l} u = x^2; \quad dv = e^{-n \frac{x-a}{\lambda}} dx \\ du = 2x dx; \quad v = -\frac{\lambda}{n} e^{-n \frac{x-a}{\lambda}} \end{array} \right| = \\ &= -x^2 e^{-n \frac{x-a}{\lambda}} \Big|_a^{+\infty} + 2 \int_a^{+\infty} x e^{-n \frac{x-a}{\lambda}} dx = a^2 + 2 \left(-\frac{\lambda}{n} x e^{-n \frac{x-a}{\lambda}} \Big|_a^{+\infty} + \frac{\lambda}{n} \int_a^{+\infty} e^{-n \frac{x-a}{\lambda}} dx \right) = \\ &= a^2 + 2 \cdot \frac{\lambda}{n} \cdot \frac{na + \lambda}{n} = a^2 + \frac{2na\lambda + 2\lambda^2}{n^2}; \\ \mathbb{D}a^* &= a^2 + \frac{2na\lambda + 2\lambda^2}{n^2} - \frac{(na + \lambda)^2}{n^2} = \frac{\lambda^2}{n^2}; \end{aligned} \tag{24}$$

$$\left\{ \begin{array}{l} \lim_{n \rightarrow \infty} \mathbb{D}a^* = a^2 + 0 - a^2 = 0; \\ a^* - \text{асимптотично незміщена.} \end{array} \right. \Rightarrow$$

$$\Rightarrow a^* - \text{консистентна} \tag{25}$$

Для дослідження на консистентність параметра λ^* краще піти через означення границі за ймовірністю, оскільки обчислення дисперсії потребує обчислення коваріації, що буде доволі проблематично.

$$\mathbb{P} \lim_{n \rightarrow \infty} \lambda^* = \mathbb{P} \lim_{n \rightarrow \infty} \bar{\xi} - a^* = \mathbb{E}\xi - \mathbb{P} \lim_{n \rightarrow \infty} \min\{\xi_i\} = a + \lambda - a = \lambda \Rightarrow$$

$$\Rightarrow \lambda^* - \text{консистентна} \tag{26}$$

Ефективність.

Оскільки оцінки виявилися зміщеними, перевіряти їх на ефективність немає сенсу.

6. Інтервальні оцінки параметрів.

Задача: побудувати інтервальні оцінки параметрів розподілу з довірчою імовірністю $\gamma = 0.95$.

Оскільки розподіл нестандартний, доволі важко буде знайти точний довірчий інтервал, але при великих n доволі непоганої точності можна досягти і з використанням асимптотичних довірчих інтервалів, тому знайдемо саме такий інтервал.

Озн. $(\theta_{1n}^*, \theta_{2n}^*)$ називається асимптотичним довірчим інтервалом для параметру θ , якщо $\lim_{n \rightarrow \infty} \mathbb{P}\{\theta_{1n}^* < \theta < \theta_{2n}^*\} \stackrel{(\geq)}{=} \gamma$.

При великих n

$$\frac{\theta^* - \mathbb{E}\theta^*}{\sqrt{\mathbb{D}\theta^*}} \approx \mathcal{N}(0, 1)$$

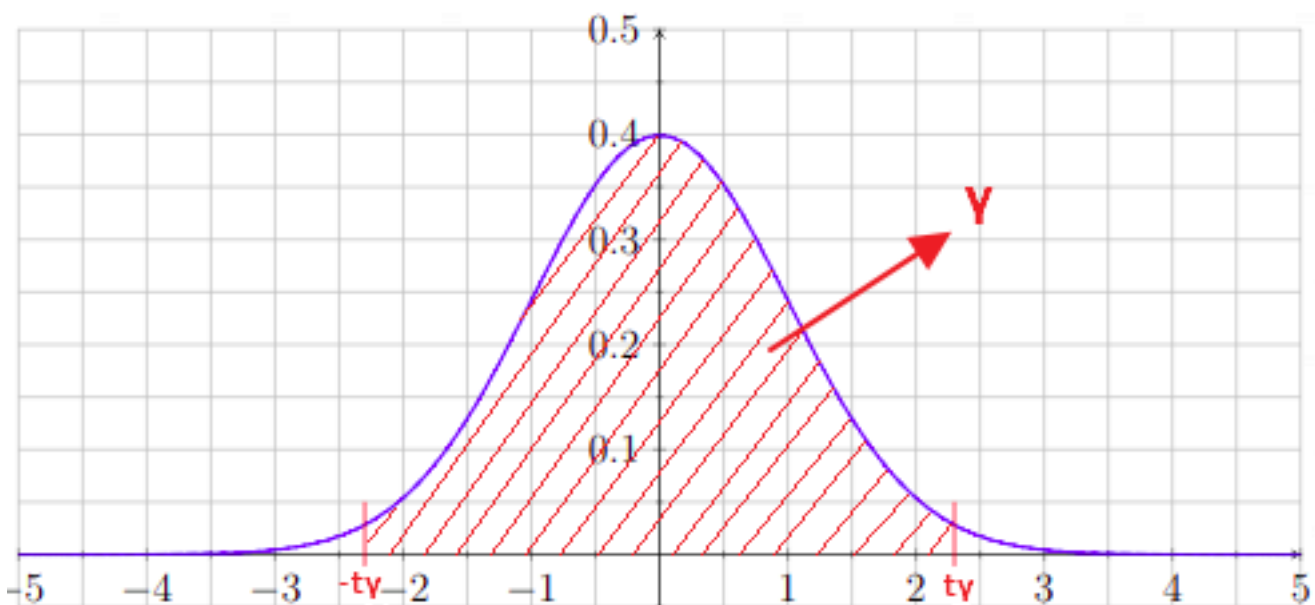


Рис. 6. Стандартний нормальний розподіл.

$$t_\gamma : \Phi(t_\gamma) = \frac{\gamma}{2} \quad (27)$$

$$\mathbb{P}\{-t_\gamma < \mathcal{N}(0, 1) < t_\gamma\} = \gamma$$

$$\mathbb{P}\{-t_\gamma < \frac{\theta^* - \mathbb{E}\theta^*}{\sqrt{\mathbb{D}\theta^*}} < t_\gamma\} = \gamma \quad (28)$$

Для знаходження асимптотичної інтервальної оцінки параметра θ потрібно розв'язати нерівність під оператором \mathbb{P} відносно θ .

Для $\gamma = 0.95$ маємо значення $t_\gamma = 1.95996$.

Асимптотичний інтервал для параметра зміщеного середнього λ .

Оскільки для даної оцінки рахувати дисперсію потрібно через коваріацію, бо її компоненти залежні, то можна спробувати трохи знизити точність, але спростити обчислення – прийняти a^* відомим, підставивши оцінене раніше обчислену оцінку ММВ.

$$\begin{aligned}
 \frac{\lambda^* - \mathbb{E}\lambda^*}{\sqrt{\mathbb{D}\lambda^*}} &= \left| \frac{\mathbb{D}\lambda^* = \mathbb{D}(\bar{\xi} - a) = \mathbb{D}\bar{\xi} = \frac{1}{n}\mathbb{D}\xi_i = \frac{1}{n}\mathbb{D}\xi =}{\sqrt{\mathbb{D}\lambda^*}} \right| = \frac{1}{n} (\mathbb{E}\xi^2 - (\mathbb{E}\xi)^2) = \frac{1}{n} ((a + \lambda)^2 + \lambda^2 - (a + \lambda)^2) = \frac{\lambda^2}{n} \left| = \right. \\
 &= \frac{\lambda^* - \frac{(n-1)\lambda}{n}}{\lambda} \sqrt{n} = \frac{n\lambda^* - (n-1)\lambda}{n\lambda} \sqrt{n} = \frac{n\lambda^* - (n-1)\lambda}{\sqrt{n}\lambda}; \\
 -t_\gamma < \frac{n\lambda^* - (n-1)\lambda}{\sqrt{n}\lambda} < t_\gamma &\Rightarrow \frac{(n\lambda^* - (n-1)\lambda)^2}{n\lambda^2} < t_\gamma^2 \Rightarrow \\
 &\Rightarrow (n-1)^2\lambda^2 - 2(n^2 - n)\lambda^*\lambda + n^2(\lambda^*)^2 < nt_\gamma^2\lambda^2 \Rightarrow \\
 &\Rightarrow (n^2 - 2n - nt_\gamma^2 + 1)\lambda^2 - 2(n^2 - n)\lambda^*\lambda + n^2(\lambda^*)^2 < 0 \quad \Rightarrow \\
 \mathcal{D} = b^2 - 4\hat{a}c &= 4(n^2 - n)^2(\lambda^*)^2 - 4(n^2 - 2n - nt_\gamma^2 + 1)n^2(\lambda^*)^2 = \\
 &= 4(\lambda^*)^2(\cancel{n^4} - \cancel{2n^3} + \cancel{n - n^4} + \cancel{2n^3} + n^3t_\gamma^2 - n^2) = 4n(\lambda^*)^2(n^2t_\gamma^2 - n + 1); \\
 \Rightarrow \lambda_{1,2}^* &= \frac{-b \pm \sqrt{\mathcal{D}}}{2\hat{a}} = \frac{2(n^2 - n)\lambda^* \pm 2\lambda^* \sqrt{n(n^2t_\gamma^2 - n + 1)}}{2(n^2 - n(2 - t_\gamma^2) + 1)} = \\
 &= \frac{\lambda^* \left((n^2 - n) \pm \sqrt{n(n^2t_\gamma^2 - n + 1)} \right)}{n^2 - n(2 + t_\gamma^2) + 1} \quad (29)
 \end{aligned}$$

Отже:

$$\lambda \in (4.45658, \quad 6.65321).$$

Як видно, отримана за ММВ оцінка потрапляє у заданий проміжок, однак оцінка методом моментів – ні.

Асимптотичний інтервал для параметра зміщення a .

$$\begin{aligned}
\frac{a^* - \mathbb{E}a^*}{\sqrt{\mathbb{D}a^*}} &= \frac{\min\{\xi_i\} - \frac{na+\lambda}{n}}{\frac{\lambda}{n}} = \frac{n}{\lambda} \min\{\xi_1\} - \frac{na}{\lambda} - 1; \\
-t_\gamma &< \frac{n}{\lambda} \min\{\xi_1\} - \frac{na}{\lambda} - 1 < t_\gamma \Rightarrow \\
\Rightarrow -t_\gamma - \frac{n}{\lambda} \min\{\xi_1\} + 1 &< -\frac{na}{\lambda} < t_\gamma - \frac{n}{\lambda} \min\{\xi_1\} + 1 \Rightarrow \\
\Rightarrow \frac{\lambda t_\gamma}{n} + \min\{\xi_i\} - \frac{\lambda}{n} &> a > -\frac{\lambda t_\gamma}{n} + \min\{\xi_i\} - \frac{\lambda}{n}; \\
a &\in \left(-\frac{\lambda t_\gamma}{n} + \min\{\xi_i\} - \frac{\lambda}{n}, \quad \frac{\lambda t_\gamma}{n} + \min\{\xi_i\} - \frac{\lambda}{n} \right) \quad (30)
\end{aligned}$$

Оскільки ми не маємо точного значення параметра λ , також можемо взяти його оцінку, що знизить точність, але інтервал і так був асимптотичним, тому це не сильно вплине. Отже:

$$\begin{aligned}
a &\in \left(-\frac{\lambda^* t_\gamma}{n} + \min\{x_i\} - \frac{\lambda}{n}, \quad \frac{\lambda^* t_\gamma}{n} + \min\{x_i\} - \frac{\lambda}{n} \right) = \\
&= (-4.6464, \quad -4.4393).
\end{aligned}$$

Як видно, отримана методом максимальної вірогідності оцінка потрапляє у заданий проміжок, однак оцінка методом моментів – ні.

7. Перевірка гіпотези про розподіл.

Була висунута складна гіпотеза про розподіл $H_0 = \{\xi \sim \text{Exp}(\lambda, a)\}$. Будемо перевіряти її за критерієм Пірсона (χ^2) з рівнем значущості $\alpha = 0.05$.

Маємо вектор $\vec{\theta}$ оцінок параметрів розподілу такий, що

$$\vec{\theta} = \arg \min_{\vec{\theta}^* \in \Theta} \{\chi^2(n; \vec{\theta}^*)\}$$

, де

$$\chi^2(n, \vec{\theta}^*) = \sum_{i=1}^m \frac{(n_i - np_i(\vec{\theta}^*))^2}{np_i(\vec{\theta}^*)} \xrightarrow{d} \chi_{m-s-1}^2$$

, де m – кількість інтервалів, у які потрапляють вибіркові дані для оцінювання, n_i – відповідні частоти потрапляння у інтервали, np_i – середня кількість потрапляння у інтервал, якби гіпотеза була вірною, s – кількість параметрів розподілу.

$$\begin{aligned} \mathbb{P}\{\xi \in (c, d) | \xi \sim \text{Exp}(\lambda, a)\} &= F_{\text{Exp}(\lambda, a)}(d) - F_{\text{Exp}(\lambda, a)}(c) = \\ &= (1 - e^{-\frac{d-a}{\lambda}}) \cdot \mathbb{I}\{d \geq a\} - (1 - e^{-\frac{c-a}{\lambda}}) \cdot \mathbb{I}\{c \geq a\} - \end{aligned}$$

Інтервал	$[-4.49, -2.575)$	$[-2.575, -0.66)$	$[-0.66, 1.255)$	$[1.255, 3.17)$
n_i	22	21	15	18
p_i	0.304016	0.21159	0.147263	0.102493
np_i	30.4016	21.159	14.7263	10.2493
Інтервал	$[3.17, 5.085)$	$[5.085, 7)$	$[7, 8.915)$	$[8.915, 10.83)$
n_i	10	8	3	2
p_i	0.0713335	0.049647	0.0345535	0.0240487
np_i	7.13335	4.9647	3.45535	2.40487
Інтервал	$[10.83, 12.745)$	$[12.745, 14.66)$	$[14.66, 16.575)$	$[16.575, +\infty)$
n_i	0	0	0	1
p_i	0.0167375	0.011649	0.00810754	0.0185606
np_i	1.67375	1.1649	0.810754	1.85606

Таблиця 3. Частоти та середні частоти для розподілу $\text{Exp}(5.2838, -4.49)$.

$$\chi^2(100; 5.2838, -4.49) = \sum_{i=1}^6 \frac{(n_i - np_i(5.2838, -4.49))^2}{np_i(5.2838, -4.49)} \approx 15.3694;$$

$$\chi^2_{12-2-1,0.05} = \chi^2_{9,0.05} \approx 16.92 > \chi^2(n; \lambda^*, a^*).$$

Як видно, отримане значення менше відповідного значення χ^2 , отже, нарівні значущості $\alpha = 0.05$ гіпотезу про розподіл приймаємо.

8. Висновки.

У даній роботі було на практиці використано отримані теоретичні знання та проведено аналіз вибірки із ГС та досліджено її розподіл. Роботу можна умовно поділити на 2 частини – обробку вибірових даних, розрахунок описових статистик та побудову відповідних графіків; пошук можливого розподілу, обґрунтування його вибору, побудову оцінок параметрів та перевірку гіпотези про розподіл.

В результаті було доведено, що розподіл є зміщеним експоненційним $Exp(\lambda, a)$ з ймовірними значеннями параметрів $\lambda = 5.2838, a = -4.49$.