Завдання.

2	0	0	1	0	0	0	2	2	2	0	0	0	0	0	3	0	3	2	0	1	1	0	0	4
3	0	2	2	5	0	6	0	5	0	11	0	2	1	0	2	2	3	4	0	0	1	4	1	1
1	0	1	0	2	2	0	0	1	2	0	1	1	0	3	0	1	0	4	2	5	1	2	0	1
1	0	0	1	4	1	1	3	0	2	7	3	0	1	0	2	1	0	2	0	0	0	2	0	2

Для конкретної реалізації вибірки виконати:

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 χ^2 (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.

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

Для початку відсортуємо нашу реалізацію вибірки:

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	3	3	3	3	3	3	3	4	4	4	4	4	5	5	5	6	7	11

Як бачимо, в вибірці досить мало унікальних значень, а саме 9, тому побудуємо дискретний варіаційний ряд:

варіанти	0	1	2	3	4	5	6	7	11
частоти n_i	41	21	20	7	5	3	1	1	1
частості ω_i	0.41	0,21	0,2	0,07	0.05	0.03	0.01	0.01	0.01

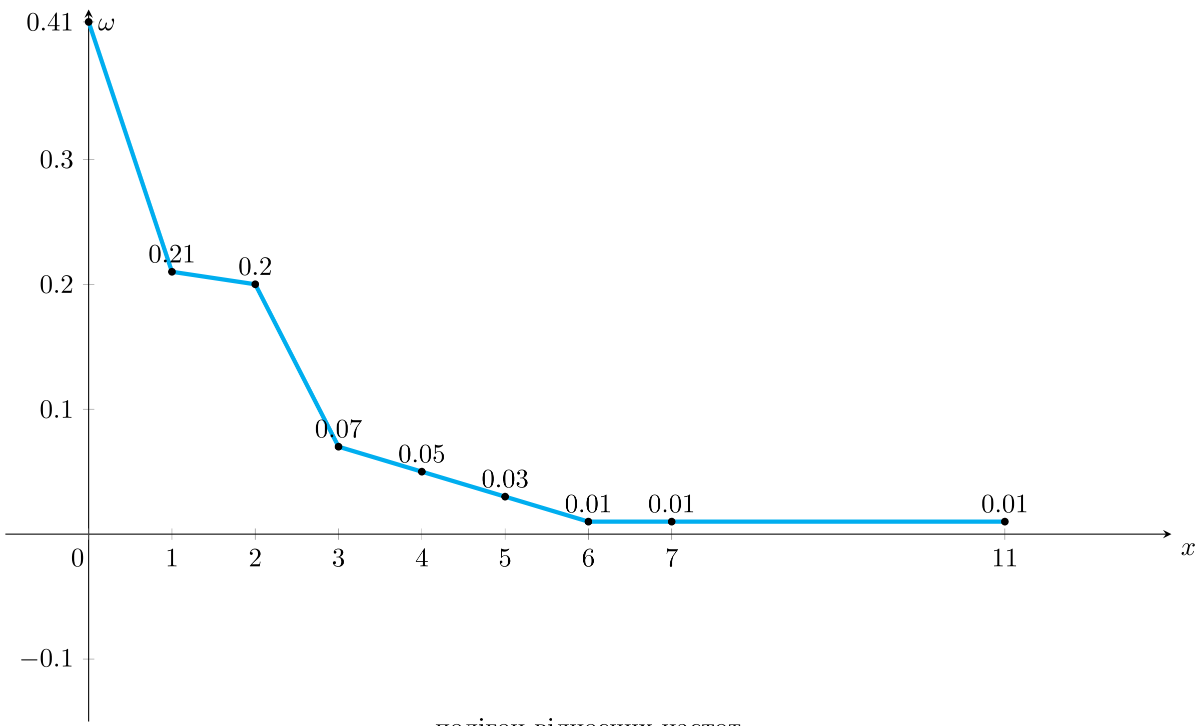
Зробимо перевірку, сумма всіх частот має дорівнювати обсягу вибірки(100) , а сума всіх частостей має дорівнювати одиниці:

$$\sum_{i=1}^{10} n_i = 41 + 21 + 20 + 7 + 5 + 3 + 1 + 1 + 1 = 100$$

$$\sum_{i=1}^{10} \omega_i = 0.41 + 0.21 + 0.2 + 0.07 + 0.05 + 0.03 + 0.01 + 0.01 + 0.01 = 1$$

2 Зробити графічне зображення вибірки.

Побудуємо за дискретним варіаційним рядом полігон відносних частот:



полігон відносних частот

Тут числа над точками це ординати відповідних точок, тобто значення частостей.

3 Побудувати емпіричну функцію розподілу

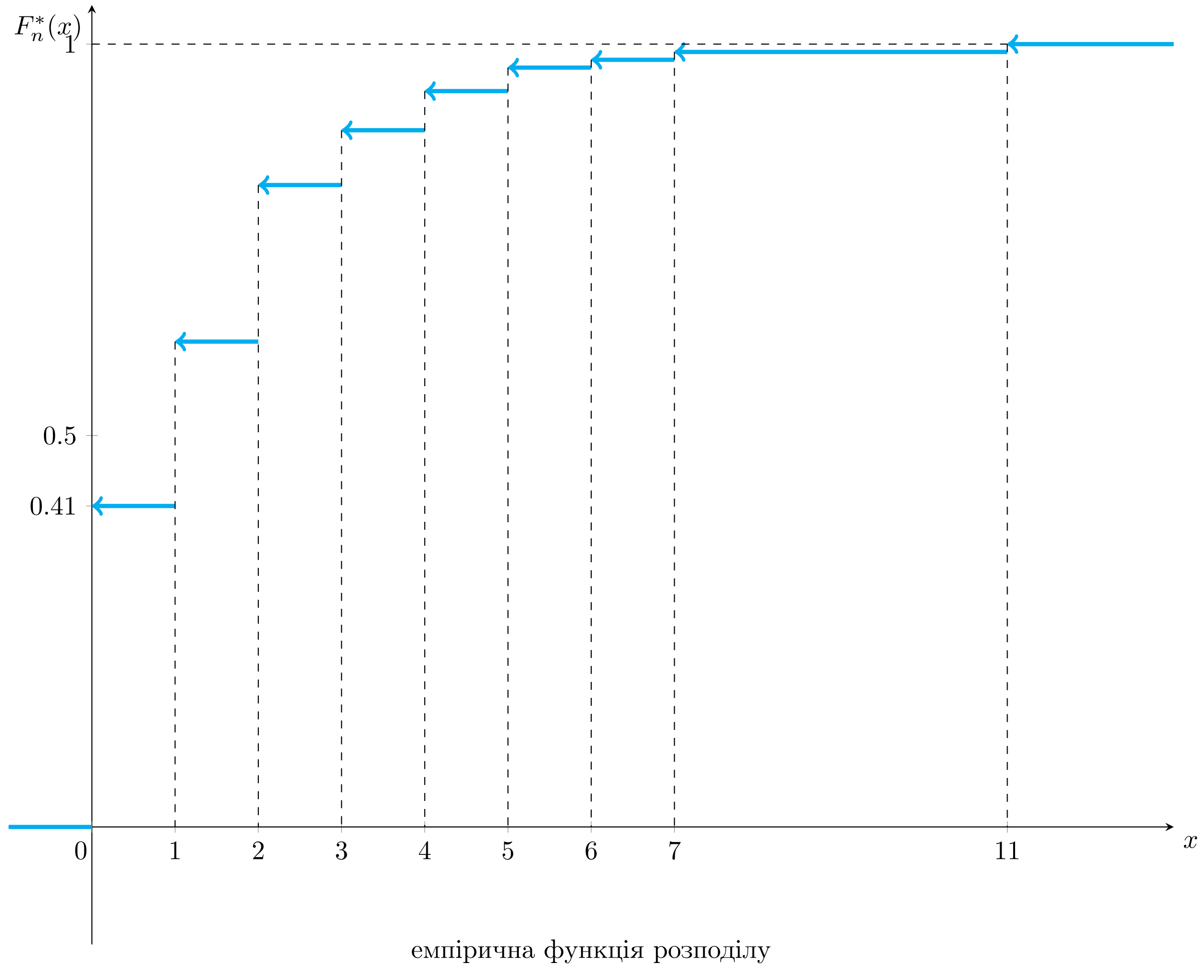
За дискретним варіаційним рядом побудуємо емпіричну функцію розподілу. Для ДВР вона визначається наступним чином:

$$F_n^*(x) = \begin{cases} 0, & x \leqslant x_1^* \\ \omega_1 = \dfrac{n_1}{n}, & x_1^* < x \leqslant x_2^* \\ \omega_2 = \dfrac{n_1 + n_2}{n}, & x_2^* < x \leqslant x_3^* \\ ... & \\ 1, & x > x_r^* \end{cases}$$

Отже маємо:

$$F_n^*(x) = \begin{cases} 0, & x \leqslant 0 \\ 0.41, & 0 < x \leqslant 1 \\ 0.41 + 0.21 = 0.62, & 1 < x \leqslant 2 \\ 0.41 + 0.21 + 0.2 = 0.82, & 2 < x \leqslant 3 \\ 0.41 + 0.21 + 02 + 0.07 = 0.89, & 3 < x \leqslant 4 \\ 0.41 + 0.21 + 02 + 0.07 + 0.05 = 0.94, & 4 < x \leqslant 5 \\ 0.41 + 0.21 + 02 + 0.07 + 0.05 + 0.03 = 0.97, & 5 < x \leqslant 6 \\ 0.41 + 0.21 + 02 + 0.07 + 0.05 + 0.03 + 0.01 = 0.98, & 6 < x \leqslant 7 \\ 0.41 + 0.21 + 02 + 0.07 + 0.05 + 0.03 + 0.01 + 0.01 = 0.99, & 7 < x \leqslant 11 \\ 1, & x > 11 \end{cases}$$

Побудуємо графік емпіричної функції розподілу



емпірична функція розподілу

4 Знайти незміщену оцінку математичного сподівання та дисперсії.

Як відомо з лекцій: незміщеною оцінкою математичного сподівання є вибіркове середнє [1]. А при невідомому математичному сподіванні незміщеною оцінкою дисперсії є виправлена вибіркова дисперсія [1]. Вибіркове середнє і виправлена вибіркова дисперсія задаються наступним чином:

$$\bar{\xi} = \frac{1}{n}\sum_{i=1}^{n} \xi_i \qquad\qquad \mathbb{D}^{**}\xi = \frac{n}{n-1}\mathbb{D}^*\xi = \frac{1}{n-1}\sum_{i=1}^{n}\left(\xi_i - \bar{\xi}\right)^2, \tag{4.1}$$

де $\mathbb{D}^*\xi$ це вибіркова дисперсія.

Для нашої реалізації вибірки знайдемо значення вибіркового середнього та виправленої вибіркової дисперсії:

$$\bar{x} = \frac{1}{100}\sum_{i=1}^{100} x_i = \frac{1}{100}(0 \cdot 41 + 1 \cdot 21 + 2 \cdot 20 + 3 \cdot 7 + 4 \cdot 5 + 5 \cdot 3 + 6 + 7 + 11) = 1.41$$

$$(\mathbb{D}^{**}\xi)_{\text{зн}} = \frac{1}{99}\sum_{i=1}^{100}(x_i - \bar{x})^2 = 41{\cdot}(0{-}1.41)^2{+}21{\cdot}(1{-}1.41)^2{+}20{\cdot}(2{-}1.41)^2{+}7{\cdot}(3{-}1.41)^2{+}...{+}(11{-}1.41)^2 = 3.29(48)$$

5 Обчислити значення вибіркової медіани, моди, асиметрії.

Знайдемо значення вибіркової медіани $(Me^*\xi)_{\text{зн}}$. У випадку побудови ДВР вона визначається як середня за розташуванням варіанта(або середнє арифметичне варіант, якщо їх парна кількість), тобто в нашому випадку маємо: $(Me^*\xi)_{\text{зн}} = x_5^* = 4$. Є ще один спосіб знайти значення вибіркової медіани, це знаходження першої варіанти, накопичена частість якої перевищила 0.5. Тоді за графіком ЕФР маємо: $(Me^*\xi)_{\text{зн}} = x_1^* = 2$

Вибіркова мода у випадку побудови ДВР визначається як варінта з найбільшою частістю, тобто у нашому випадку: $(Mo^*\xi)_{\text{зн}} = x_1^* = 0$.

Вибіркова асиметрія визначається наступною формулою:

$$As^*\xi = \frac{\frac{1}{n}\sum\limits_{i=1}^{n}\left(\xi_i - \bar{\xi}\right)^3}{\left(\mathbb{D}^*\xi\right)^{3/2}}$$

Для знаходження значення вибіркової асиметрії знайдемо значення вибіркової дисперсії. З формули 4.1, виправленої вибіркової дисперсії отримаємо:

$$(\mathbb{D}^*\xi)_{\text{зн}} = \frac{n-1}{n}(\mathbb{D}^{**}\xi)_{\text{зн}} = \frac{99}{100} \cdot 3.29(48) = 3.2619$$

Тепер можемо знайти значення вибркової асиметрії:

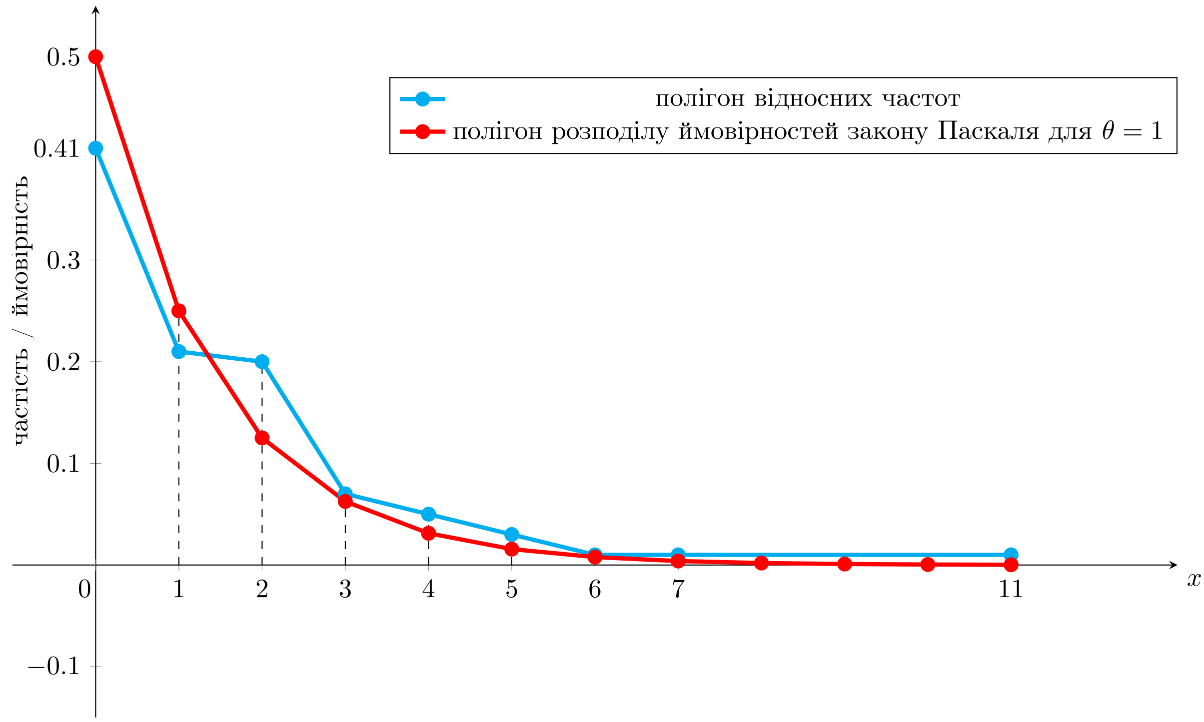
$$(As^*\xi)_{\text{зн}} = \frac{\frac{1}{100}\sum\limits_{i=1}^{100}(x_i - \bar{x})^3}{(\mathbb{D}^*\xi)_{\text{зн}}^{3/2}} = \frac{12.9489}{(3.2619)^{3/2}} = 2.198$$

6 Висунути гіпотезу про розподіл, за яким отримано вибірку.

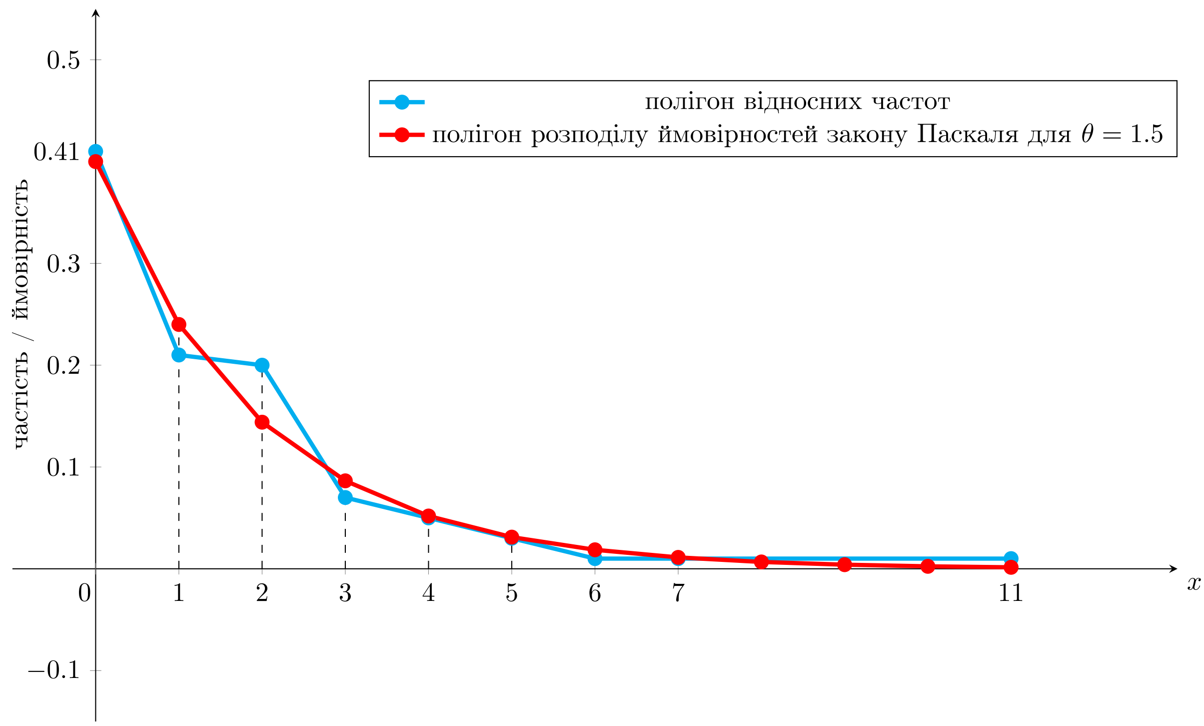
Розглянемо розподіл Паскаля. Нагадаємо, ДВВ ζ розподілена за Законом Паскаля, якщо набуває значень $0, 1, 2, \ldots$ з ймовірностями:

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

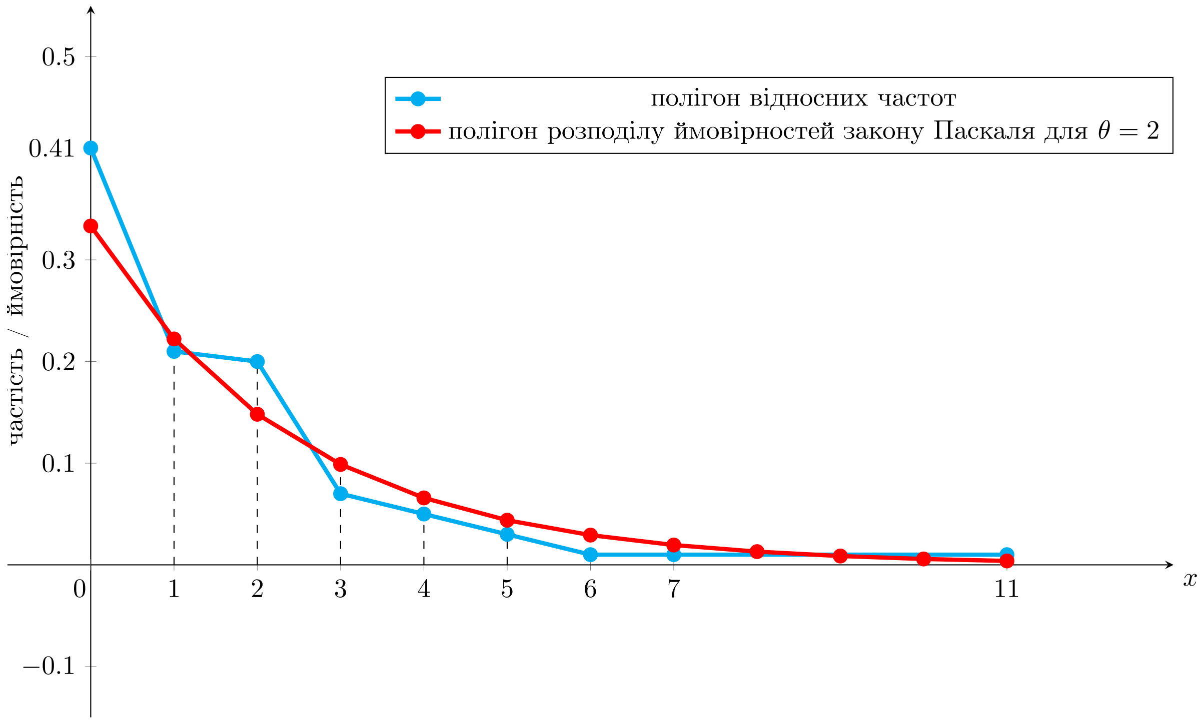
1. Розглянемо полігон розподілу ймовірностей закону Паскаля для різних параметрів



полігон відносних частот та полігон розподілу ймовірностей закону Паскаля для $\theta = 1$



полігон відносних частот та полігон розподілу ймовірностей закону Паскаля для $\theta = 1.5$

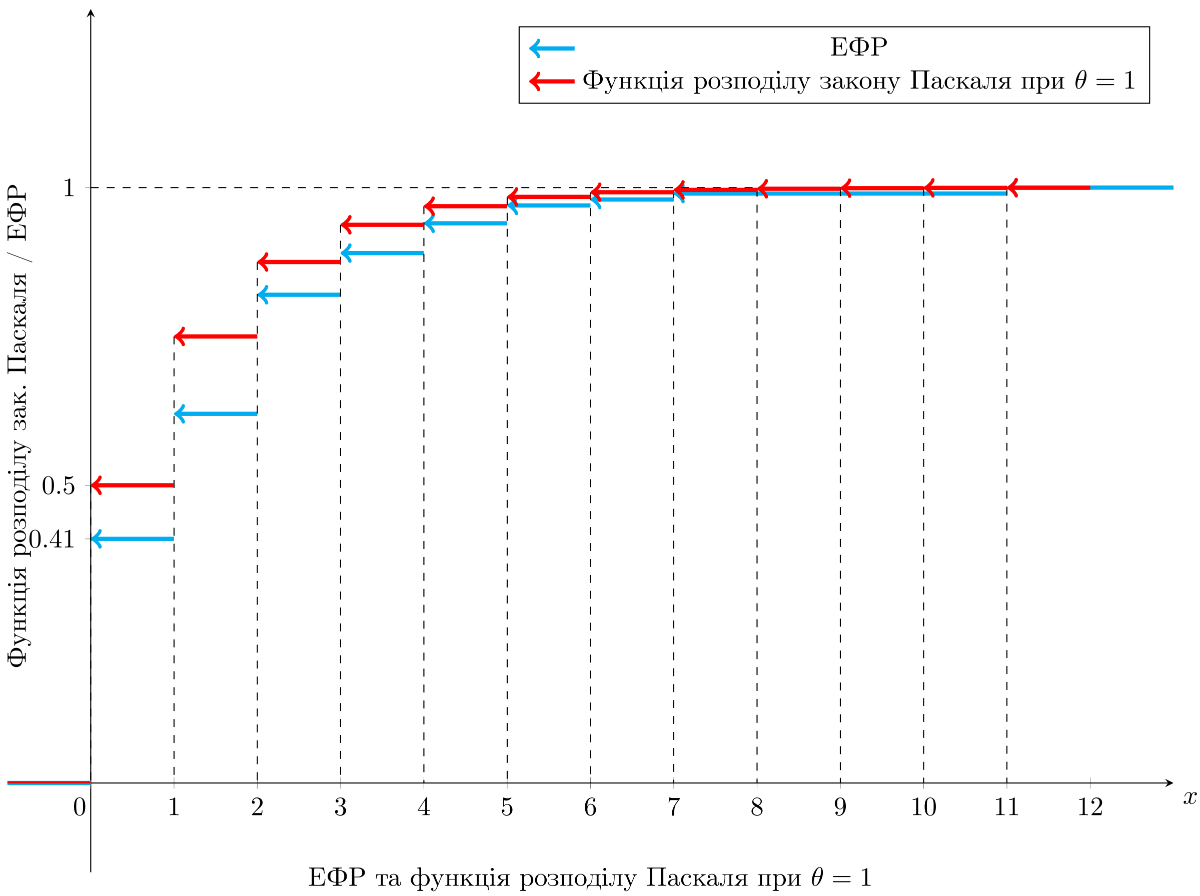


полігон відносних частот та полігон розподілу ймовірностей закону Паскаля для $\theta = 2$

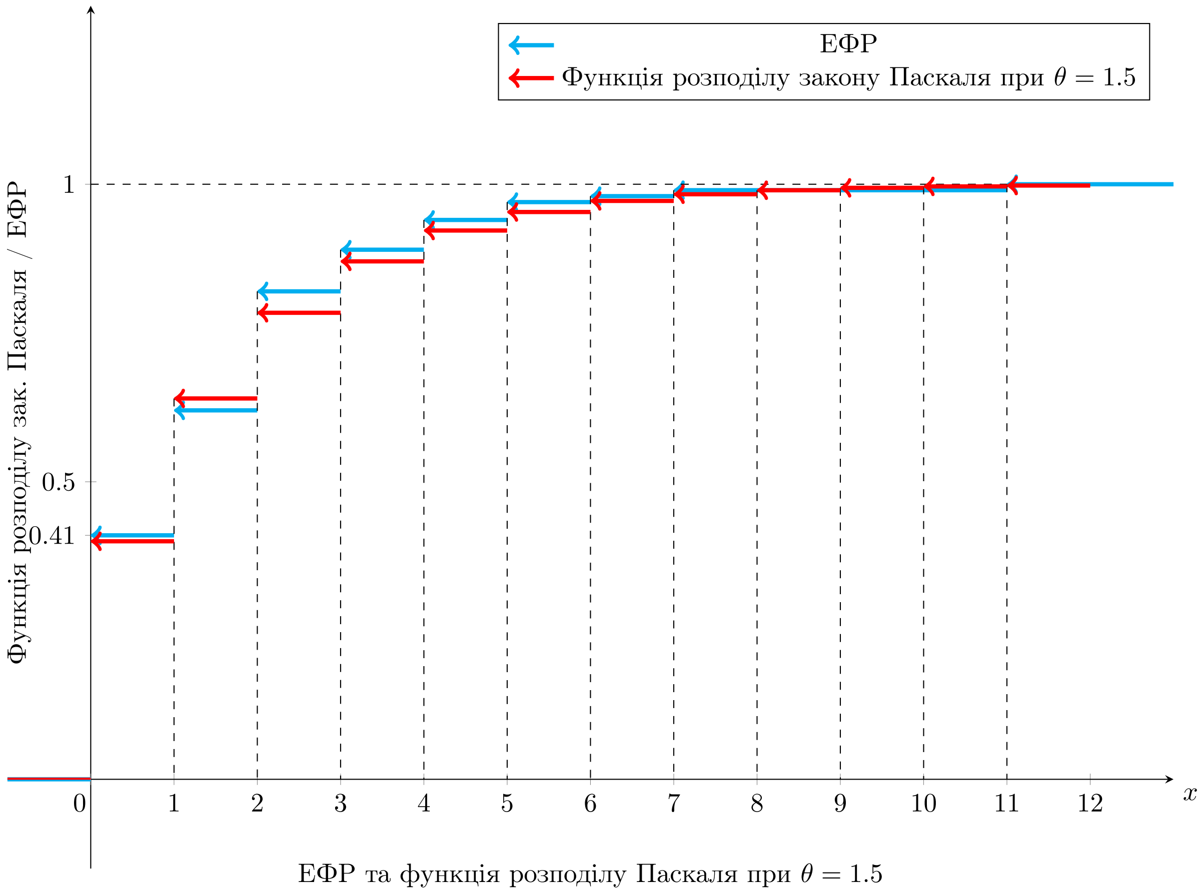
Зауваження Червоні графіки не перетинають вісь абсцис. Це просто дефект відображення LaTeX.

Отже, з графіків видно, що при певних значеннях параметра θ(а саме 1, 1.5, 2) закону Паскаля його полігон розподілу схожий на полігон відносних частот нашої реалізації вибірки.

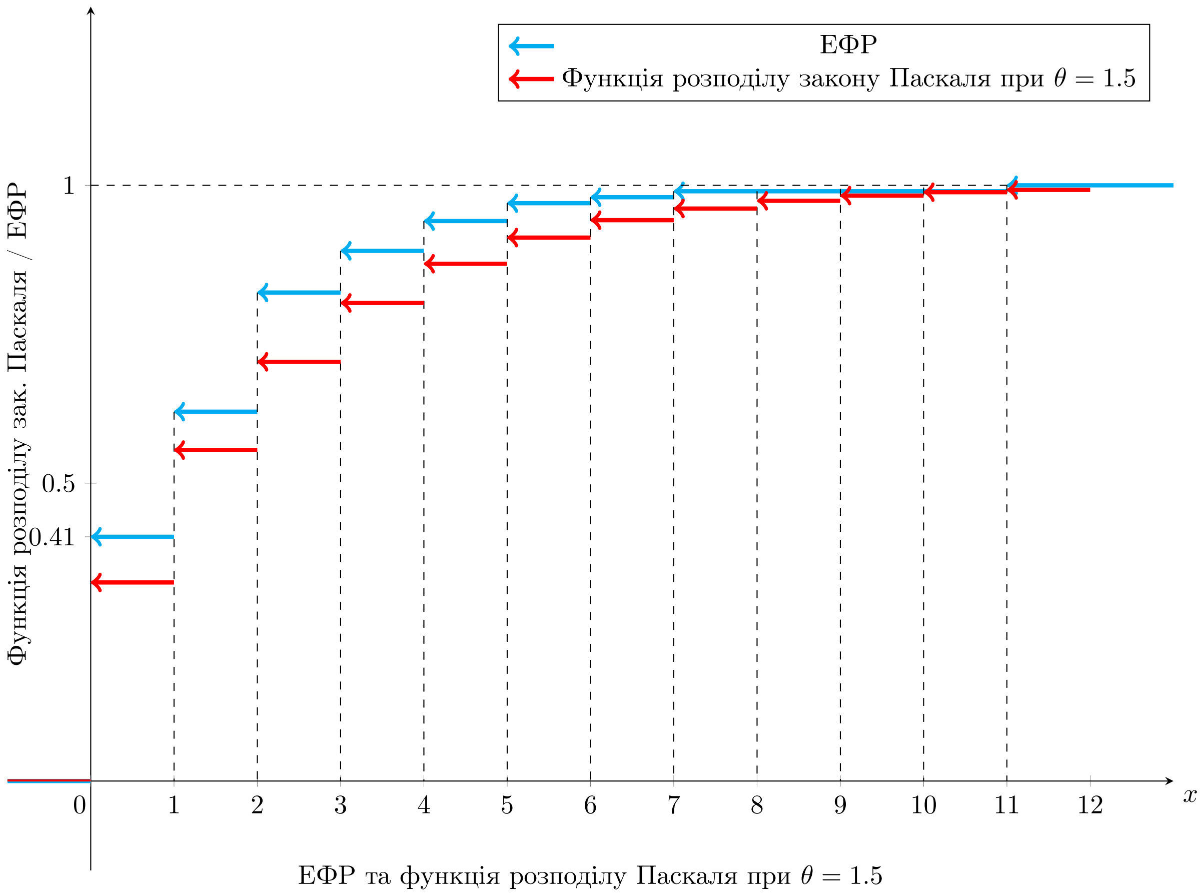
2. Розглянемо графіки функцій розподілу Паскаля при різних значеннях параметра θ



ЕФР та функція розподілу Паскаля при $\theta = 1$



ЕФР та функція розподілу Паскаля при $\theta = 1.5$



ЕФР та функція розподілу Паскаля при $\theta = 1.5$

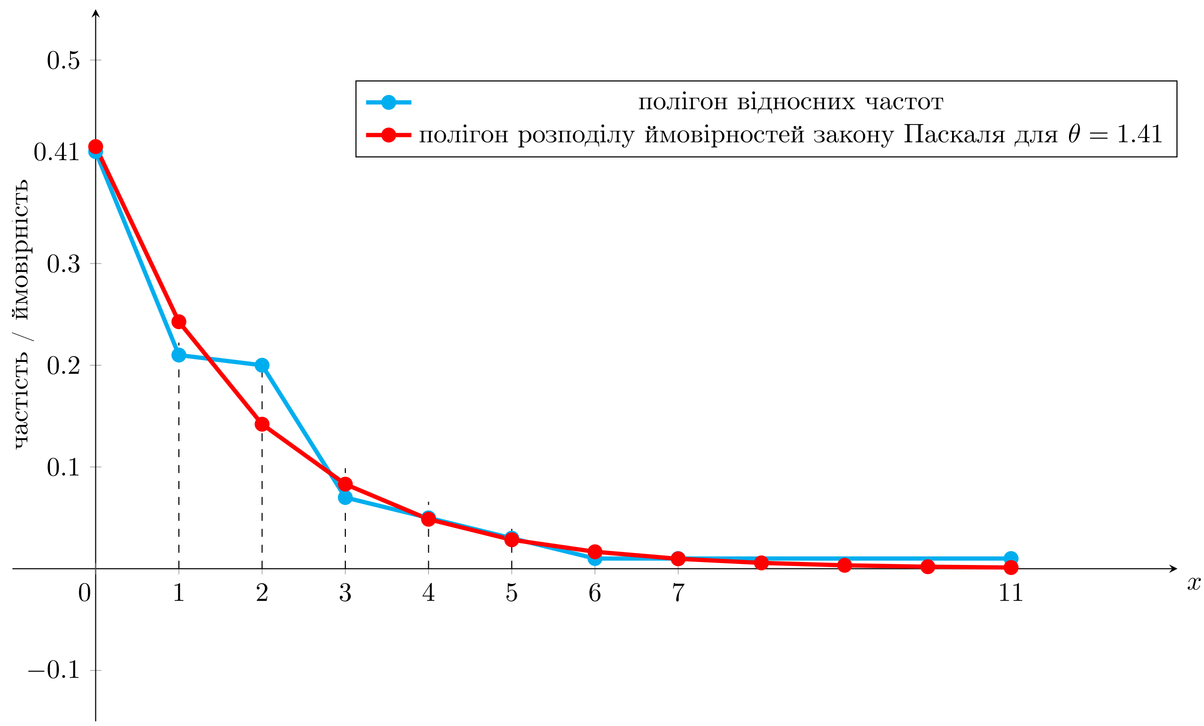
Зауваження Червоні графіки не дотикаються до лінії $y = 1$. Це також дефект відображення LaTeX

Як бачимо, графік емпіричної функції розподілу закона Паскаля при певних параметрах θ(а саме при 1, 1.5, 2) схожий на графік емпіричної функції розподілу реалізації вибірки.

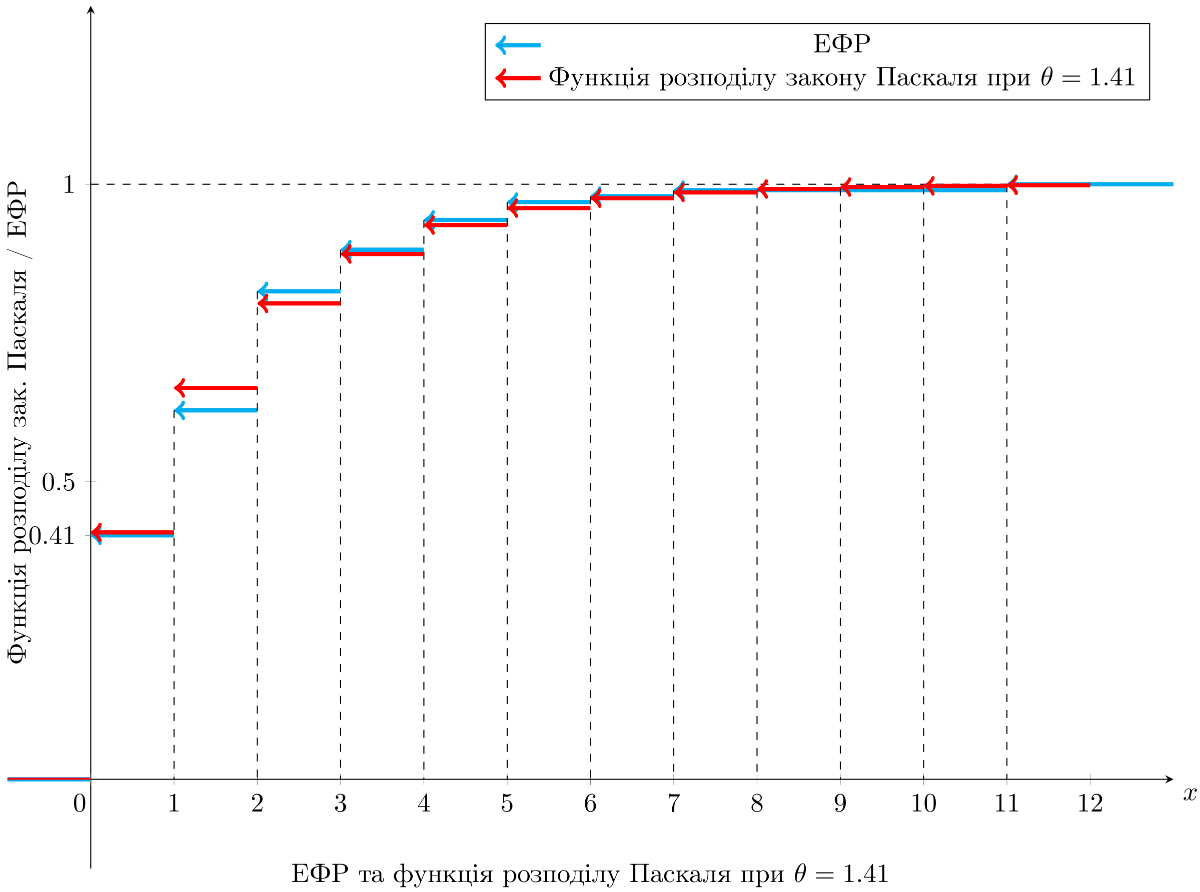
3. При будь-яких значеннях параметра θ мода випадкової величини розподіленої за законом Паскаля дорівнює нулю, в той же час значення вибіркової моди для нашої реалізації дорівнює нулю.

4. Відомо що, якщо $\zeta \sim Pas(\theta)$, то $\mathbb{E}\zeta = \theta, \mathbb{D}\zeta = \theta + \theta^2$. Якщо замість параметру θ взяти значення вибіркового середнього(далі буде показано, що вибіркове середнє це незміщена, консистентна та ефективна точкова оцінка невідомого параметру θ) , то величини $(\mathbb{D}^{**}\xi)_{\text{зн}}$ і $\bar{x} + (\bar{x})^2$ будуть досить схожими(3.29(48) і 3.3981 відповідно)

5. Також, якщо взяти заміть парметру θ значення вибіркового середнього, то полігон відносних частот і полігон розподілу ймовірностей закону Паскаля будуть досить схожими. Також при такому значенні параметра будуть дуже схожі графіки ЕФР і функція розподілу закону Паскаля.(див. наступні графіки)



полігон відносних частот та полігон розподілу ймовірностей закону Паскаля для $\theta = 1.41$



ЕФР та функція розподілу Паскаля при $\theta = 1.41$

Отже, підсумуємо, що отримали

- Полігон відносних частот реалізації вибірки схожий на полігон розподілу ймовірностей закону Паскаля.(див. ст. 5-6)
- Графік емпіричної функції розподілу нашої реалізації вибірки схожий на графік фунції розподілу закону Паскаля (див. ст. 6-7)
- При будь-яких значеннях параметра θ мода випадкової величини розподіленої за законом Паскаля дорівнює нулю, в той же час значення вибіркової моди для нашої реалізації дорівнює нулю.
- Відомо що, якщо $\zeta \sim Pas\,(\theta)$, то $\mathbb{E}\zeta = \theta, \mathbb{D}\zeta = \theta + \theta^2$. Якщо замість параметру θ взяти значення вибіркового середнього(далі буде показано, що вибіркове середнє це незміщена, конзистентна та ефективна точкова оцінка невідомого параметру θ) , то величини $(\mathbb{D}^{**}\xi)_{\text{зн}}$ і $\bar{x} + (\bar{x})^2$ будуть досить схожими(3.29(48) і 3.3981 відповідно)
- Також, якщо взяти заміть парметру θ значення вибіркового середнього, то полігон відносних частот і полігон розподілу ймовірностей закону Паскаля будуть досить схожими. Також при такому значенні параметра будуть дуже схожі графіки ЕФР і функція розподілу закону Паскаля. (див. ст. 8)

Таким чином, на основі вище написаних тверджень мною висувається гіпотеза, що генеральна сукупність, якою породжена данна вибірка, розподілена за законом Паскаля.

7 Знайти точкову оцінку параметрів гіпотетичного закону розподілу та перевірити їх властивості.

Знайдемо точкові оцінки невідомого параметра θ двома способами:

1. Метод моментів:

   Відомо, що для $\xi \sim Pas(\theta) : \mathbb{E}\xi = \theta$. Замінивши початковий момент першого порядку на емпіричний початковий момент першого порядку отримаємо рівняння Пірсона:

$$(\theta^*)_{\text{мм}} = \mathbb{E}^*\xi \quad \Rightarrow \quad (\theta^*)_{\text{мм}} = \frac{1}{n}\sum_{i=1}^{n}\xi_i = \bar{\xi}$$

2. Метод максимальної правдоподібності:

   Спочатку знайдемо функцію правдоподібності закону Паскаля:

$$\mathcal{L}(\vec{x};\theta) = \prod_{k=1}^{n}\mathbb{P}\{\xi = x_k\} = \prod_{k=1}^{n}\frac{1}{1+\theta}\left(\frac{\theta}{1+\theta}\right)^{x_k} = \left(\frac{1}{1+\theta}\right)^{n}\left(\frac{\theta}{1+\theta}\right)^{\sum\limits_{k=1}^{n}x_k} = \frac{\theta^{\sum\limits_{k=1}^{n}x_k}}{(1+\theta)^{n+\sum\limits_{k=1}^{n}x_k}}$$

   Тепер знайдемо логарифмічну функцію правдоподібності закону Паскаля:

$$\ln\mathcal{L}(\vec{x};\theta) = \ln\frac{\theta^{\sum\limits_{k=1}^{n}x_k}}{(1+\theta)^{n+\sum\limits_{k=1}^{n}x_k}} = \ln\theta\cdot\sum_{k=1}^{n}x_k - \ln(1+\theta)\cdot\left(n+\sum_{k=1}^{n}x_k\right)$$

   Тепер знайдемо частинну похідну $\dfrac{\partial\ln\mathcal{L}(\vec{x};\theta)}{\partial\theta}$ і прирівняємо її до нуля(таким чином отримаємо рівняння правдоподібності):

$$\frac{\partial\ln\mathcal{L}(\vec{x};\theta)}{\partial\theta} = \frac{1}{\theta}\sum_{k=1}^{n}x_k - \frac{\left(n+\sum\limits_{k=1}^{n}x_k\right)}{1+\theta} = \frac{\sum\limits_{k=1}^{n}x_k + \theta\cdot\sum\limits_{k=1}^{n}x_k - \theta\cdot n - \theta\cdot\sum\limits_{k=1}^{n}x_k}{\theta(1+\theta)} = \frac{-\theta\cdot n + \sum\limits_{k=1}^{n}x_k}{\theta(1+\theta)} = 0$$

   Тоді:

$$-\theta\cdot n + \sum_{k=1}^{n}x_k = \quad\Rightarrow\quad \theta = \frac{1}{n}\sum_{k=1}^{n}x_k$$

   Отримали $\theta_{\text{кр}} = \frac{1}{n}\sum\limits_{k=1}^{n}x_k$ - критична точка функції $\ln\mathcal{L}(\vec{x};\theta)$. Тепер перевіримо чи є ця точка саме максимумом цієї функції. Для цього має виконуватись умова:

$$\left.\frac{\partial^2\ln\mathcal{L}(\vec{x};\theta)}{\partial\theta^2}\right|_{\theta=\frac{1}{n}\sum\limits_{k=1}^{n}x_k} < 0$$

маємо:

$$\frac{\partial^2 \ln \mathcal{L}(\vec{x};\theta)}{\partial \theta^2} = \frac{-n\cdot(\theta+\theta^2)-(1+2\theta)\cdot\left(-\theta\cdot n+\sum\limits_{k=1}^{n} x_k\right)}{\theta^2(1+\theta)^2} = \frac{-n\cdot\theta^2-\sum\limits_{k=1}^{n} x_k+2n\cdot\theta^2-2\theta\sum\limits_{k=1}^{n} x_k}{\theta^2(1+\theta)^2}$$

$$\left.\frac{\partial^2 \ln \mathcal{L}(\vec{x};\theta)}{\partial \theta^2}\right|_{\theta=\frac{1}{n}\sum\limits_{k=1}^{n} x_k} = \frac{-\frac{1}{n}\left(\sum\limits_{k=1}^{n} x_k\right)^2-\sum\limits_{k=1}^{n} x_k+\frac{2}{n}\left(\sum\limits_{k=1}^{n} x_k\right)^2-\frac{2}{n}\left(\sum\limits_{k=1}^{n} x_k\right)^2}{\left(\frac{1}{n}\sum\limits_{k=1}^{n} x_k\right)^2\cdot\left(1+\frac{1}{n}\sum\limits_{k=1}^{n} x_k\right)^2} = \frac{-\left(\frac{1}{n}\left(\sum\limits_{k=1}^{n} x_k\right)^2+\sum\limits_{k=1}^{n} x_k\right)}{\left(\frac{1}{n}\sum\limits_{k=1}^{n} x_k\right)^2\cdot\left(1+\frac{1}{n}\sum\limits_{k=1}^{n} x_k\right)^2}$$

Оскільки $\sum\limits_{k=1}^{n} x_k > 0, \quad n > 0$, то :

$$\left.\frac{\partial^2 \ln \mathcal{L}(\vec{x};\theta)}{\partial \theta^2}\right|_{\theta=\frac{1}{n}\sum\limits_{k=1}^{n} x_k} < 0$$

Отримали, що $\theta_{\text{кр}} = \frac{1}{n}\sum\limits_{k=1}^{n} x_k$ - точка максимуму функції $\ln \mathcal{L}(\vec{x};\theta)$. Отже, отримали точкову оцінку невідомого параметру θ розподілу Паскаля за методом максимальної правдоподібності:

$$(\theta^*)_{\text{ммп}} = \sum_{k=1}^{n} \xi_k = \bar{\xi}$$

Обома методами отримали однакові точкові оцінки невідомого параметру θ :

$$\theta^* = (\theta^*)_{\text{мм}} = (\theta^*)_{\text{ммп}} = \frac{1}{n}\sum_{k=1}^{n} \xi_k = \bar{\xi}$$

Перевіримо властивості точкової оцінки θ^* :

1. **Незміщеність**

$$\mathbb{E}\theta^* = \mathbb{E}\frac{1}{n}\sum_{k=1}^{n}\xi_k = \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\xi_k = \left|\text{всі } \xi_k \text{ розподілені як } \xi\right| = \frac{1}{n}\cdot n\cdot\mathbb{E}\xi = \theta$$

Отже, точкова оцінка θ^* - незміщена

2. **Консистентність**

$\{\xi_k\}_{k=1}^{n}$ - незалежні та однаково розподілені випадкові величини. Для них існують математичні сподівання $\mathbb{E}\xi_k = \theta$ та дисперсії $\mathbb{D}\xi_k = \theta + \theta^2$. При чому всі дисперсії - рівномірно обмежені. Тоді за Законом Великих Чисел у формі Чебишова маємо:

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k \xrightarrow{\text{P}} \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\xi_k, \quad n\to\infty \quad \Rightarrow \quad \left|\frac{1}{n}\sum_{k=1}^{n} = \theta^*; \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}\xi_k = \theta\right| \quad \Rightarrow \quad \theta^* \xrightarrow{\text{P}} \theta, \quad n\to\infty$$

Отже, точкова оцінка - консистентна. До речі, за Посиленим законом Великих Чисел точкова оцінка θ^* є навіть сильно консистентною, тобто $\theta^* \xrightarrow{\text{P1}} \theta, n\to\infty$.

3. **Ефективність**

   На ст. 10 Отримали значення виразу:

$$\frac{\partial \ln \mathcal{L}(\vec{x};\theta)}{\partial \theta} = \frac{-\theta \cdot n + \sum\limits_{k=1}^{n} x_k}{\theta(1+\theta)}$$

   Тепер перейдемо до випадкової вибірки:

$$\frac{\partial \ln \mathcal{L}(\vec{\xi};\theta)}{\partial \theta} = \frac{-\theta \cdot n + \sum\limits_{k=1}^{n} \xi_k}{\theta(1+\theta)} = \frac{1}{\theta(1+\theta)} \cdot \frac{1}{n}(\theta^* - \theta) = C(n,\theta) \cdot (\theta^* - \theta)$$

   Отже, за наслідком з нерівності Рао-Крамера оцінка $\theta^* = \bar{\xi}$ є ефективною.

4. **Асимптотична нормальність**

$$\frac{\theta^* - \theta}{\sqrt{\mathbb{D}\theta^*}} = \frac{\frac{1}{n}\sum\limits_{k=1}^{n} \xi_k - \frac{1}{n} \cdot n \cdot \mathbb{E}\xi}{\sqrt{\mathbb{D}\frac{1}{n}\sum\limits_{k=1}^{n} \xi_k}} = \left| \text{всі } \xi_k \text{ незалежні та однаково розподілені} \right| = \frac{\frac{1}{n}\sum\limits_{k=1}^{n} (\xi_k - \mathbb{E}\xi_k)}{\sqrt{\frac{1}{n^2} \cdot \sum\limits_{k=1}^{n} \mathbb{D}\xi_k}} =$$

$$= \frac{\sum\limits_{k=1}^{n} (\xi_k - \mathbb{E}\xi_k)}{\sum\limits_{k=1}^{n} \sigma_k^2} \xrightarrow{F} \nu \sim N(0,1), n \to \infty$$

   Останній граничний перехід був записаний за наслідком до теореми Ляпунова: "Якщо всі ξ_k однаково розподілені, то умова Ляпунова виконується автоматично".

Підсумуємо. В цьому розділі ми знайшли точкову оцінку невідомого параметру закону Паскаля: $\theta^* = \bar{\xi}$, а також довели, що ця точкова оцінка є незміщеною, консистентною, ефективною, а також асимптотично нормальною.

8 Перевірити за допомогою критерію χ^2 (Пірсона) гіпотезу про розподіл з рівнем значущості $\alpha = 0.05$.

Критерій χ^2 опирається на теорему Пірсона. Вона формулюється так:

Теорема 1 *(теорема Пірсона)*

1. *Якщо* ***проста*** *гіпотеза H_0 щодо закону розподілу ГС справджується, то статистика η прямує за розподілом до розподілу χ^2_{r-1} (розподіл χ^2 з $r-1$ ступенями вільності) при $n \to \infty$*

2. *Якщо* ***складна*** *гіпотеза H_0 щодо закону розподілу ГС справджується, то статистика η прямує за розподілом до розподілу χ^2_{r-s-1}(розподіл χ^2 з $r-s-1$ ступенями вільності) при $n \to \infty$. Тут s - кількість невідомих параметрів розподілу, які оцінюються.*

Статистика η визначається як:

$$\eta = \sum_{i=1}^{r} \frac{n}{p_i}\left(\frac{n_i}{n} - p_i\right)^2 = \sum_{i=1}^{r} \frac{(n_i - np_i)^2}{np_i}$$

Згідно з висновками розділу 6 висувається гіпотеза $H_0 : \xi \sim Pas(1.41)$. За нашою гіпотезою генеральна сукупність може приймати такі значення $X = \{0, 1, 2, 3 \dots\}$. Розіб'ємо множину X на r підмножин за таким правилом:

1. Якщо $r \geqslant 20$, то потрібно, щоб виконувалсь $np_i \geqslant 5$
2. Якщо $r < 20$, то потрібно, щоб виконувалсь $np_i \geqslant 20$

Отже, маємо таке розбиття X:

$$X_0 = \{0\}; \quad X_1 = \{1\}; \quad X_2 = \{2\}; \quad X_3 = \{3, 4, 5 \dots\}$$

Обчислимо ймовірності $p_i = \mathbb{P}\{\xi \in X/H_0\}$. тут n_i це кількість значень з реалізації вибірки, які попали в множину X_i.

X_i	$\{0\}$	$\{1\}$	$\{2\}$	{3, 4 , 5 ...}
p_i	0.41494	0.24276	0.14203	0.20027
np_i	41.494	24.276	14.203	20.027
n_i	41	21	20	18

Виконаємо перевірку для значень p_i:

$$\sum_{i=1}^{4} p_i = 0.41494 + 0.24276 + 0.14203 + 0.20027 = 1$$

Тепер обчислимо значення статистики η взятої з теореми Пірсона:

$$\eta_{\text{зн}} = \frac{(41 - 41.494)^2}{41.494} + \frac{(21 - 24.276)^2}{24.276} + \frac{(20 - 14.203)^2}{14.203} + \frac{(18 - 20.027)^2}{20.027} \approx 3.0191$$

Тепер за таблицею розподілу Пірсона знайдемо значення $t_{\text{кр}}$. $r - s - 1 = 4 - 1 - 1 = 2$, тому оскільки $\alpha = 0.05$ маємо: $t_{\text{кр}} = 5.99$. Отримали $\eta_{\text{кр}} < t_{\text{кр}}$, отже, на рівні значущості 0.05 дані не суперечать висунутій гіпотезі про те, що генеральна сукупність розподілена за законом Паскаля($\xi \sim Pas(1.41)$).

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

Побудуємо довірчий інтерал такий, що: $\mathbb{P}\left\{|\theta^* - \theta| < \varepsilon\right\} = \gamma = 0.95$. В розділі 6 було доведено, що точкова оцінка $\theta^* = \bar{\xi}$ є асимптотично нормальною, тому розподіл $\frac{\theta^* - \theta}{\sqrt{\mathbb{D}\theta}}$ можна вважати приблизно $N(0, 1)$. Тоді довірчий інтервал можна знайти з наступної рівності:

$$\mathbb{P}\left\{\frac{|\theta^* - \theta|}{\sqrt{\mathbb{D}\theta^*}} < t_\gamma\right\} \approx 2\Phi\left(t_\gamma\right)$$

Знайдемо $\mathbb{D}\theta^*$:

$$\mathbb{D}\theta^* = \mathbb{D}\frac{1}{n}\sum_{k=1}^{n}\xi_k = \left|\text{всі } \xi_k \text{ - незалежні та однаково розподілені}\right| = \frac{1}{n^2} \cdot n \cdot \mathbb{D}\xi = \frac{1}{n} \cdot \left(\theta + \theta^2\right)$$

За допомогою таблиці значень функції Лапласа знайдемо таке t_γ, щоб $2\Phi(t_\gamma) \approx 0.95: \quad t_\gamma \approx 1.44$

Тепер розв'яжемо нерівність $\dfrac{|\theta^* - \theta|}{\sqrt{\mathbb{D}\theta^*}} < t_\gamma$ відносно θ:

$$\frac{|\theta^* - \theta|}{\sqrt{\mathbb{D}\theta^*}} < t_\gamma \quad \Rightarrow \quad \frac{(\theta^* - \theta)^2}{\mathbb{D}\theta^*} < t_\gamma^2 \quad \Rightarrow \quad (\theta^* - \theta)^2 < t_\gamma^2 \cdot \frac{1}{n} \cdot \left(\theta + \theta^2\right)$$

$$(\theta^*)^2 - 2 \cdot \theta \cdot \theta^* + \theta^2 < \frac{t_\gamma^2}{n} \cdot \theta + \frac{t_\gamma^2}{n} \cdot \theta^2$$

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

$$D = \left(2 \cdot \theta^* + \frac{t_\gamma^2}{n}\right)^2 - 4 \cdot \left(1 - \frac{t_\gamma^2}{n}\right) \cdot (\theta^*)^2 = 4 \cdot \theta^* \cdot \frac{t_\gamma^2}{n} + \frac{t_\gamma^4}{n^2} + \frac{t_\gamma^2}{n} \cdot (\theta^*)^2$$

Отримали такий довірчий інтервал з рівнем надійності γ:

$$\left(\frac{2 \cdot \theta^* + \frac{t_\gamma^2}{n} - \sqrt{4 \cdot \theta^* \cdot \frac{t_\gamma^2}{n} + \frac{t_\gamma^4}{n^2} + \frac{t_\gamma^2}{n} \cdot (\theta^*)^2}}{2 \cdot \left(1 - \frac{t_\gamma^2}{n}\right)}; \quad \frac{2 \cdot \theta^* + \frac{t_\gamma^2}{n} + \sqrt{4 \cdot \theta^* \cdot \frac{t_\gamma^2}{n} + \frac{t_\gamma^4}{n^2} + \frac{t_\gamma^2}{n} \cdot (\theta^*)^2}}{2 \cdot \left(1 - \frac{t_\gamma^2}{n}\right)} \right)$$

Підставивши значення отримаємо, що $\theta \in (1.1792; 1.722)$ з ймовірністю 0.95

10 Висновки.

Під час виконання данної розрахункової роботи було проведено первинну обробку деякої реалізації вибірки: було побудовано дискретний варіаційний ряд, полігон відносних частот, а також емпіричну функцію розподілу. Було знайдено значення незміщених оцінок математичного сподівання та дисперсії. Було обчислено значення деяких вибіркових характеристик генеральної сукупності, а саме вибіркову моду, вибіркову медіану, а також вибіркову асиметрію. Були порівняні полігони відносних частостей і полігони розподілу закону Паскаля при певних параметрах, а також графіки емпіричної функції розподілу і функції розподілу закону Паскаля. Згодом була висута гіпотеза про те, що генеральна сукупність, якою була отримана реалізація вибірки розподілена за законом Паскаля. Була знайдена точкова оцінка невідомого параметра, а також доведена її незміщеність, конзистнентність, ефективність і асимптотична нормальність. За допомогою криткрію χ^2(Пірсона) було показано, що на рівні значущості 0.05 дані не суперечать висунутій гіпотезі про те, що генеральна сукупність розподілена за законом Паскаля з параметром $\theta = 1.41$. В кінці було отримано довірчий інтервал з рівнем надійності $\gamma = 0.95$ для параметра θ, а також обчислено його межі для нашої реалізації вибірки.

Додаток

[1] - Каніовська І.Ю. Конспект з теорії ймовірностей та математичної статистики ст. 93