

НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
"ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ"
НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ УКРАЇНИ
"КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМЕНІ ІГОРЯ
СІКОРСЬКОГО"
КАФЕДРА МАТЕМАТИЧНИХ МЕТОДІВ СИСТЕМНОГО АНАЛІЗУ

РОЗРАХУНКОВА РОБОТА
з предмету "Математична статистика"

Виконав студент групи
КА-81 Фордуй Нікіта
Перевірила Каніовська І.Ю.

Київ 2020

1 Завдання

Дана конкретна реалізація вибірки об'ємом $n = 100$:

2 0 8 0 15 1 1 1 7 1 0 0 3 1 1 1 0 0 3 1
 2 4 10 6 1 0 1 0 0 2 0 1 5 0 1 9 4 2 11 3
 2 0 8 1 6 3 0 1 1 4 0 9 5 3 3 0 0 10 2 0
 3 11 0 9 0 1 4 1 0 2 0 1 1 3 4 7 1 3 3 0
 4 7 6 0 3 0 1 15 11 1 2 4 0 2 0 0 0 26 4 0

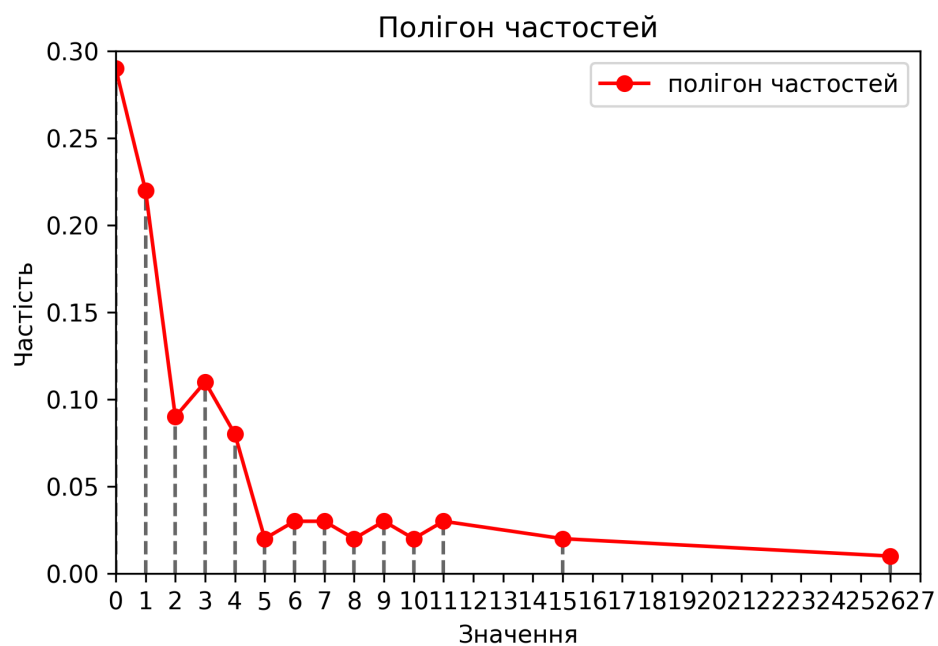
2 Побудова варіаційного ряду вибірки

Маємо невелику кількість різних значень - тому побудуємо дискретний варіаційний ряд. Підрахувавши кількість варіант (14) та їх частоти і знаючи об'єм вибірки отримаємо дискретний варіаційний ряд :

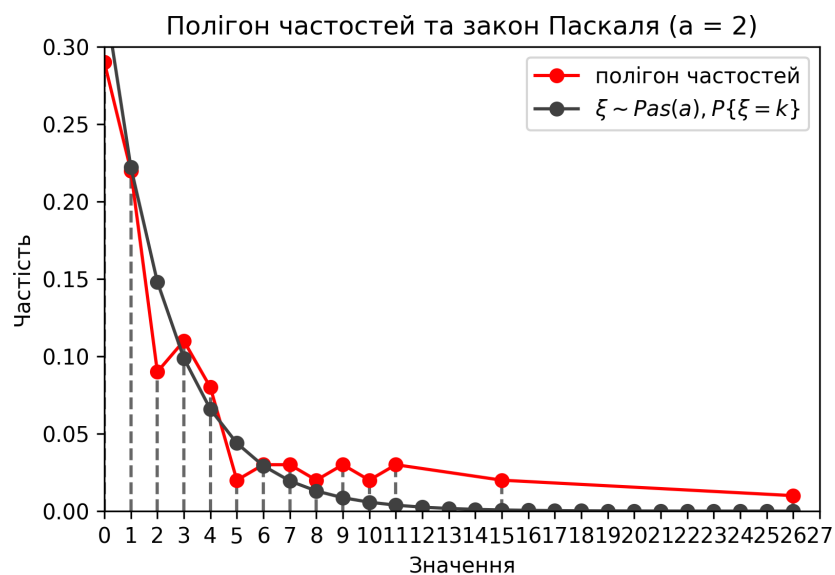
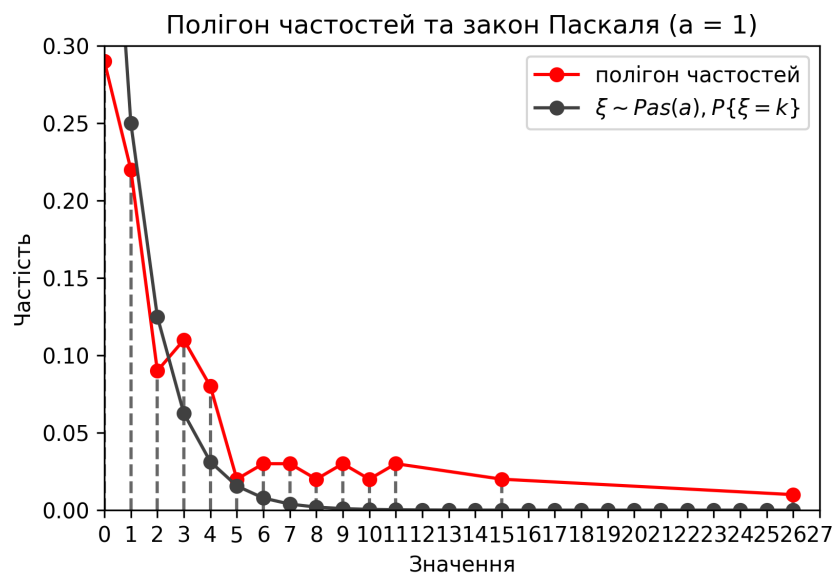
x_i^*	0	1	2	3	4	5	6	7	8	9	10	11	15	26
n_i	29	22	9	11	8	2	3	3	2	3	2	3	2	1
$\omega_i = \frac{n_i}{n}$	$\frac{29}{100}$	$\frac{22}{100}$	$\frac{9}{100}$	$\frac{11}{100}$	$\frac{8}{100}$	$\frac{2}{100}$	$\frac{3}{100}$	$\frac{3}{100}$	$\frac{2}{100}$	$\frac{3}{100}$	$\frac{2}{100}$	$\frac{3}{100}$	$\frac{2}{100}$	$\frac{1}{100}$
ω_i^H	$\frac{29}{100}$	$\frac{51}{100}$	$\frac{60}{100}$	$\frac{71}{100}$	$\frac{79}{100}$	$\frac{81}{100}$	$\frac{84}{100}$	$\frac{87}{100}$	$\frac{89}{100}$	$\frac{92}{100}$	$\frac{94}{100}$	$\frac{97}{100}$	$\frac{99}{100}$	1

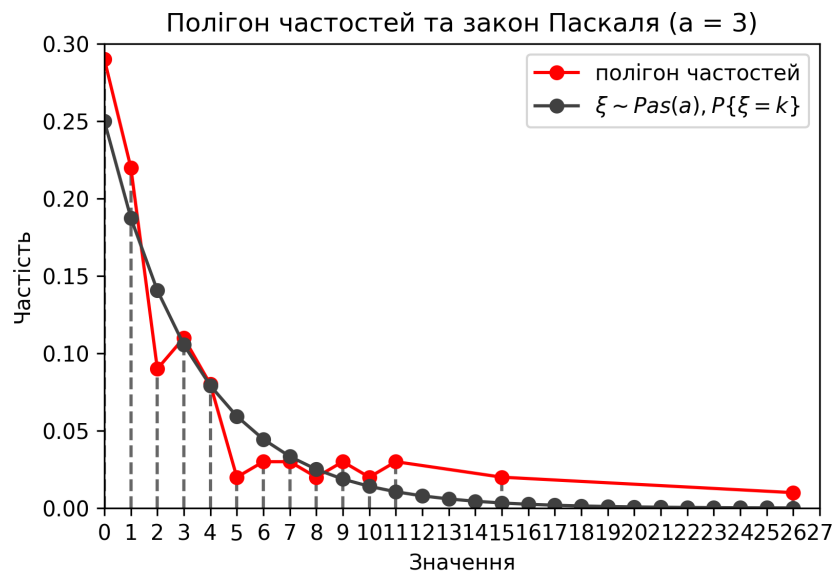
де x_i^* - варіанти реалізації вибірки, n_i - частота варіанти, $\omega_i = \frac{n_i}{n}$ - частість варіанти або відносна частота, ω_i^H - накопичена частість.

За дискретним варіаційним рядом побудуємо його геометричну інтерпретацію - полігон відносних частот (частостей):



Порівняємо полігон частостей нашої реалізації виборки із полігоном ймовірностей закону Паскаля при різних значеннях його параметра.





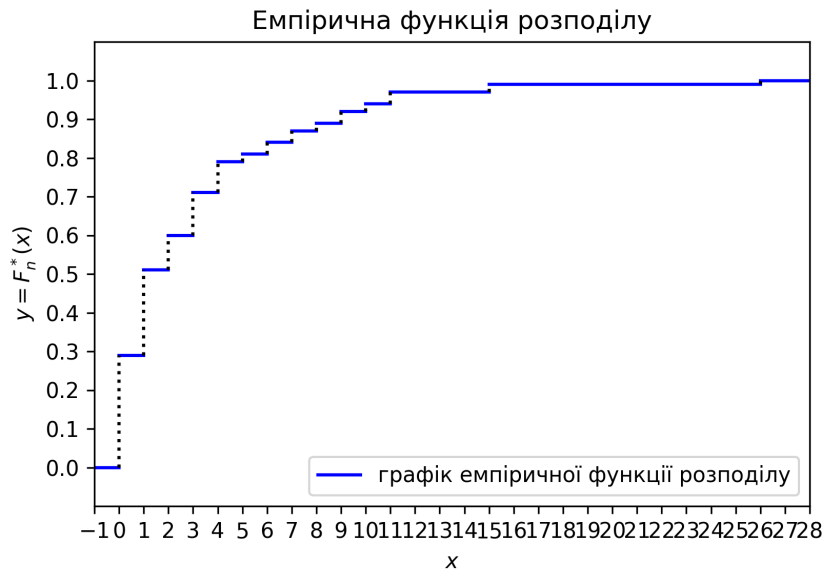
Можна побачити, що полігон ймовірностей закону Паскаля при певних значеннях його параметра ($a = 1, 2, 3, 4$) дуже схожий на полігон частотей нашої реалізації вибірки.

3 Емпірична функція розподілу

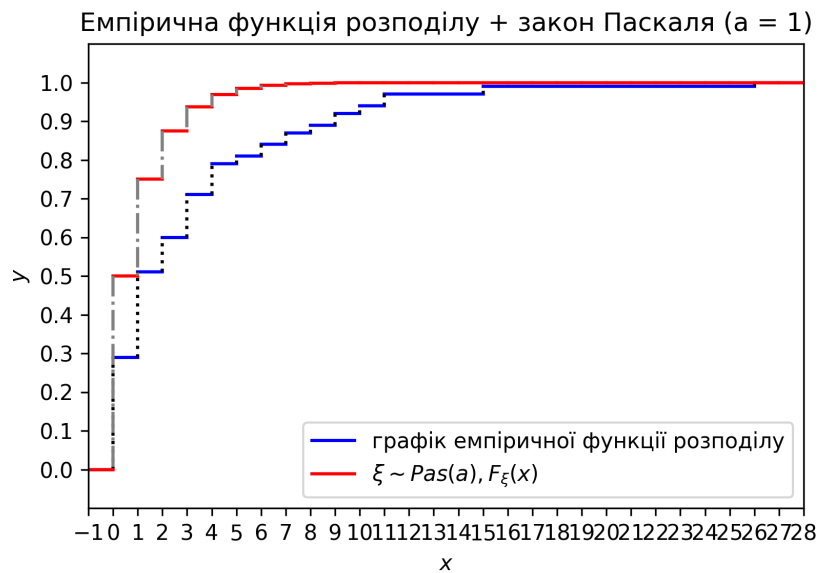
Побудуємо емпіричну функцію розподілу за вже побудованим дискретним варіаційним рядом:

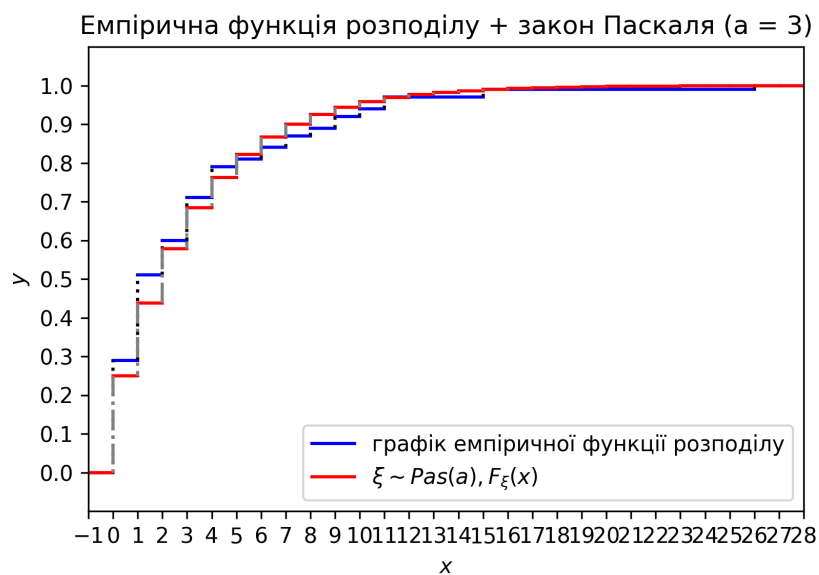
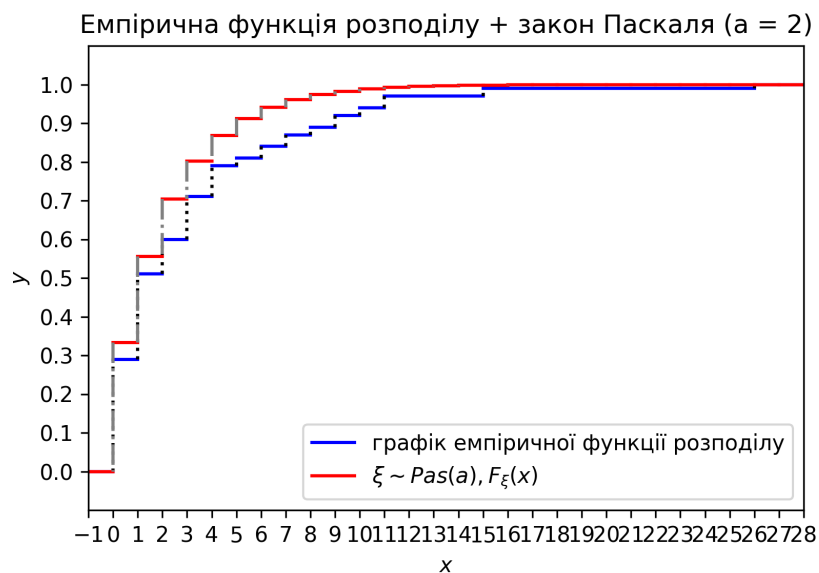
$$F_n^*(x) = \begin{cases} 0, & x \leq 0 \\ \frac{29}{100}, & 0 < x \leq 1 \\ \frac{29+22}{100} = \frac{51}{100}, & 1 < x \leq 2 \\ \frac{51}{100} + \frac{9}{100} = \frac{60}{100}, & 2 < x \leq 3 \\ \frac{60}{100} + \frac{11}{100} = \frac{71}{100}, & 3 < x \leq 4 \\ \frac{71}{100} + \frac{8}{100} = \frac{79}{100}, & 4 < x \leq 5 \\ \frac{79}{100} + \frac{2}{100} = \frac{81}{100}, & 5 < x \leq 6 \\ \frac{81}{100} + \frac{3}{100} = \frac{84}{100}, & 6 < x \leq 7 \\ \frac{84}{100} + \frac{3}{100} = \frac{87}{100}, & 7 < x \leq 8 \\ \frac{87}{100} + \frac{2}{100} = \frac{89}{100}, & 8 < x \leq 9 \\ \frac{89}{100} + \frac{3}{100} = \frac{92}{100}, & 9 < x \leq 10 \\ \frac{92}{100} + \frac{2}{100} = \frac{94}{100}, & 10 < x \leq 11 \\ \frac{94}{100} + \frac{3}{100} = \frac{97}{100}, & 11 < x \leq 15 \\ \frac{97}{100} + \frac{2}{100} = \frac{99}{100}, & 15 < x \leq 26 \\ \frac{99}{100} + \frac{1}{100} = 1, & x > 26 \end{cases} \quad (1)$$

Зобразимо емпіричну функцію розподілу геометрично :



Порівняємо графік емпіричної функції розподілу варіаційного ряду з графіком функції розподілу закону Паскаля при різних параметрах ($a = 1, 2, 3$):





З рисунків вище можна побачити що графік емпіричної функції розподілу нашої реалізації вибірки схожий при певних значеннях параметра a на функцію розподілу закону Паскаля.

4 Обчислення вибірових характеристик генеральної сукупності (медіана, мода, асиметрія)

Для початку знайдемо $(Mo_{\xi}^*)_{\text{знач.}}$ - значення вибіркової моди (тієї варіанти, якій відповідає найбільша частість). Для знаходження цієї варіанти використаємо вже побудований дискретний варіаційний ряд (див. ст. 1 пункт 2). Проаналізувавши варіаційний ряд побачимо, що:

$$(Mo_{\xi}^*)_{\text{знач.}} = x_1^* = 0$$

Зауважимо, що випадкова величина μ , розподілена за законом Паскаля при будь-яких значеннях параметра a має моду $Mo_{\mu} = 0$.

Знайдемо значення вибіркової медіани $(Me_{\xi}^*)_{\text{знач.}}$ для нашої реалізації виборки. З варіаційного ряду (див. ст. 1 пункт 2), враховуючи те, що кількість варіант - парна, знайдемо:

$$(Me_{\xi}^*)_{\text{знач.}} = \frac{x_7^* + x_8^*}{2} = 6.5$$

Для знаходження значення вибіркової асиметрії спочатку потрібно знайти значення вибіркової дисперсії, а тому й вибіркового середнього:

$$\bar{x} = (E_{\xi}^*)_{\text{знач.}} = \frac{1}{100} \sum_{k=1}^{14} x_k^* n_k = 3.06$$

За допомогою цього знайдемо значення вибіркової дисперсії:

$$(D_{\xi}^*)_{\text{знач.}} = \frac{1}{100} \sum_{k=1}^{100} (x_k - 3.06)^2 = 17.136400000000002$$

Отримавши значення вибіркової дисперсії, можна отримати значення вибіркової асиметрії для даної реалізації вибірки:

$$(As_{\xi}^*)_{\text{знач.}} = \frac{\frac{1}{100} \sum_{k=1}^{100} (x_k - 3.06)^3}{(17.136400000000002)^{\frac{3}{2}}} = 2.504088773053977$$

5 Незміщені оцінки математичного сподівання та дисперсії

ξ - генеральна сукупність, $\vec{\xi} = (\xi_1, \xi_2, \dots, \xi_n)$ - випадкова вибірка, $n = 100$ - об'єм вибірки.

За точкову оцінку математичного сподівання візьмемо вибіркове середнє:

$$E_{\xi}^* = \frac{1}{n} \sum_{k=1}^n \xi_k$$

Перевіримо незміщеність цієї точкової оцінки:

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

Відповідно, ця точкова оцінка математичного сподівання є незміщеною.

За точкову оцінку дисперсії візьмемо:

$$\begin{aligned} D_{\xi}^* &= \frac{1}{n} \sum_{k=1}^n (\xi_k - \bar{\xi})^2 = \frac{1}{n} \sum_{k=1}^n ((\xi_k - E_{\xi}) - (\bar{\xi} - E_{\xi}))^2 = \\ &= \frac{1}{n} \sum_{k=1}^n (\xi_k - E_{\xi})^2 - 2(\bar{\xi} - E_{\xi}) \frac{1}{n} \sum_{k=1}^n (\xi_k - E_{\xi}) + (\bar{\xi} - E_{\xi})^2 = \\ &= \frac{1}{n} \sum_{k=1}^n (\xi_k - E_{\xi})^2 - (\bar{\xi} - E_{\xi})^2 \end{aligned}$$

Порахуємо матсподівання цієї оцінки:

$$E(D_{\xi}^*) = \frac{1}{n} \sum_{k=1}^n E(\xi_k - E_{\xi})^2 - E(\bar{\xi} - E_{\xi})^2 = D_{\xi} - D_{\bar{\xi}}$$

Бачимо, що ця оцінка - зміщена. Знайдемо $D_{\bar{\xi}} = \frac{1}{n} \sum_{k=1}^n D_{\xi_k} = \frac{D_{\xi}}{n}$.

$$E(D_{\xi}^*) = D_{\xi} - \frac{D_{\xi}}{n} = \frac{n-1}{n} D_{\xi}$$

Тоді оцінка $D_{\xi}^{**} = \frac{n}{n-1} D_{\xi}^*$ буде незміщеною оцінкою дисперсії.

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

Обчислимо значення цих точкових оцінок на данній реалізації виборки:

$$(E_{\xi}^*)_{\text{знач.}} = \frac{1}{100} \sum_{k=1}^{14} x_k^* n_k = 3.06$$

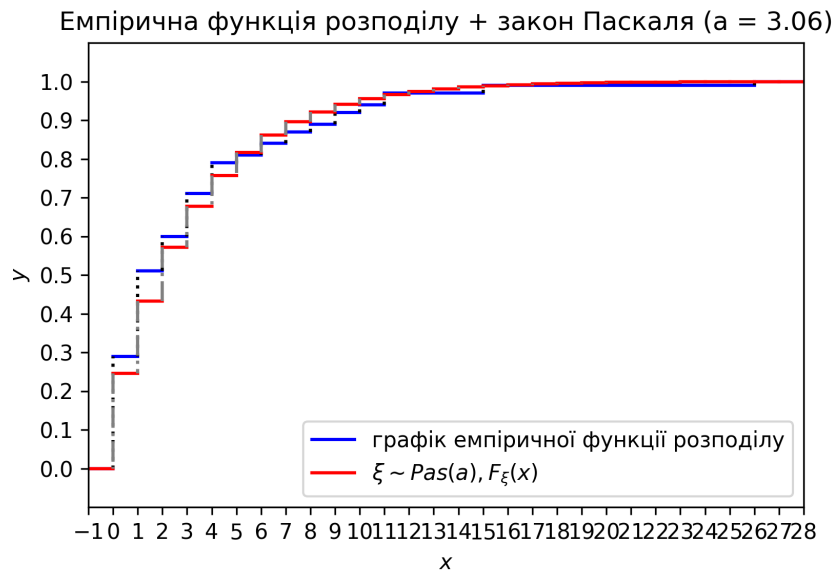
де x_k^* - к-та варіанта, n_k - частота вибірки.

$$(D_{\xi}^{**})_{\text{знач.}} = \frac{1}{99} \sum_{k=1}^{100} (x_k - 3.06)^2 = 17.309494949495$$

6 Гіпотеза про розподіл, за яким отримано вибірку

Виходячи з того що:

- полігон частотей реалізації вибірки схожий на полігон ймовірностей закону Паскаля (див. с. 3-4)
- емпірична функцію розподілу реалізації вибірки схожа на функцію розподілу закону Паскаля (див. с. 6-7)
- мода випадкової величини, розподіленої за законом Паскаля дорівнює 0 при будь-яких параметрах закону; в той же самий час значення вибіркової медіани для нашої реалізації виборки також дорівнює 0.
- в наступному параграфі буде показано, що оцінка математичного сподівання закону Паскаля $E_{\xi}^* = \frac{1}{n} \sum_{k=1}^n \xi_k$ є не тільки незміщеною, а й конзистентною та ефективною. Тоді, якщо порівняти полігон частотей даної вибірки та полігон ймовірностей, графік емпіричної функції розподілу та графік функції розподілу закону Паскаля з відповідним параметром, то вони будуть дуже схожі (див. наступні рис.)



Таким чином висувається гіпотеза, що генеральна сукупність, якою породжена данна вибірка, розподілена за законом Паскаля.

7 Точкові оцінки параметру гіпотетичного закону розподілу

Спочатку скористаємось методом моментів для знаходження точкової оцінки параметра а закону Паскаля ($\mu \sim Pas(a)$). Прирівняємо емпіричний початковий момент 1-го порядку та математичне сподівання випадкової величини, розподіленої за законом Паскаля; отримаємо рівняння Пірсона:

$$\begin{aligned} E_\mu &= E_\mu^* \\ E_\mu &= a, E_\mu^* = \frac{1}{n} \sum_{k=1}^n \xi_k \\ (a^*)_{\text{MM}} &= \frac{1}{n} \sum_{k=1}^n \xi_k \end{aligned} \quad (3)$$

Отримали статистику - точкову оцінку параметра а закону Паскаля.

Тепер отримаємо точкову оцінку параметра а за допомогою методу максимальної правдоподібності (Фішера). Спочатку знайдемо функцію правдоподібності закону Паскаля:

$$\begin{aligned} \mathcal{L}(\vec{x}, a) &= \prod_{k=1}^n \mathbb{P}\{\xi = x_k\} = \prod_{k=1}^n \frac{a^{x_k}}{(1+a)^{x_k+1}} = \frac{a^{\sum_{k=1}^n x_k}}{(1+a)^{\sum_{k=1}^n x_k + n}} \\ \ln \mathcal{L}(\vec{x}, a) &= \left(\sum_{k=1}^n x_k \right) \ln a - \left(\left(\sum_{k=1}^n x_k \right) + n \right) \ln(1+a) \\ \frac{\partial \ln \mathcal{L}(\vec{x}, a)}{\partial a} &= \frac{1}{a(1+a)} \sum_{k=1}^n x_k - \frac{n}{1+a} = 0 \end{aligned} \quad (4)$$

$$\frac{1}{a(1+a)} \sum_{k=1}^n x_k = \frac{n}{1+a}$$

$$a = \frac{1}{n} \sum_{k=1}^n x_k$$

Отримали оцінку параметра а закону Паскаля методом максимальної правдоподібності:

$$(a^*)_{\text{МП}} = \frac{1}{n} \sum_{k=1}^n \xi_k \quad (5)$$

Перевіримо виконання достатньої умови:

$$\frac{\partial^2 \ln \mathcal{L}(\vec{x}, a)}{\partial^2 a} = -\frac{1+2a}{a^2(1+a)^2} \sum_{k=1}^n x_k + \frac{n}{(1+a)^2}$$

$$\left. \frac{\partial^2 \ln \mathcal{L}(\vec{x}, a)}{\partial^2 a} \right|_{a=a^*} = -\frac{n^2}{(1+a^*)^2} \left(\frac{1}{a^*} + 1 \right) < 0 \quad (6)$$

Достатня умова виконана.

Обома методами отримали однакову оцінку параметра а:

$$(a^*)_{\text{ММ}} = (a^*)_{\text{МП}} = a^* = \frac{1}{n} \sum_{k=1}^n \xi_k = \bar{\xi}$$

Перевіримо властивості цієї оцінки:

1. **Незміщенність.** Вже доведена раніше (див. 2 стор.9)
2. **Конзистентність.** Так як $\{\xi_k\}$ - i.i.d¹, для $\forall \xi_k : E_{\xi_k} = a < \infty, D_{\xi_k} = a^2 + a$ – рівномірно обмежені, то за ЗВЧ $a^* \xrightarrow[n \rightarrow \infty]{\mathbb{P}} E_{a^*} = a$. А це і означає що оцінка конзистентна.

¹i.i.d. - Independent and identically distributed random variables - незалежні та однаково розподілені випадкові величини

3. **Ефективність.** Розглянемо вираз $\frac{\partial \ln \mathcal{L}(\vec{x}, a)}{\partial a}$. Його вже було знайдено раніше (див. 4 стор. 12).

$$\begin{aligned} \frac{\partial \ln \mathcal{L}(\vec{x}, a)}{\partial a} &= \frac{1}{a(1+a)} \sum_{k=1}^n x_k - \frac{n}{1+a} = \\ &= \frac{n}{a(1+a)} \left(\frac{1}{n} \sum_{k=1}^n x_k - a \right) = \\ &= C(n, a)(a^* - a) \end{aligned}$$

Таким чином, за наслідком з нерівності Рао-Крамера, оцінка $a^* = \frac{1}{n} \sum_{k=1}^n \xi_k$ є ефективною.

Таким чином, маємо незміщенну, конзистентну та ефективну оцінку параметра a закона Паскаля.

8 Перевірка гіпотези про розподіл

Перевірка гіпотези про розподіл генеральної сукупності буде здійснюватись за допомогою критерію χ^2 (Пірсона) з рівнем значущості $\alpha = 0.05$.

Висунемо гіпотезу $H_0 : \xi \sim Pas(3.06)$. Згідно нашої гіпотези, генеральна сукупність може приймати такі значення: $\{0, 1, 2, \dots\}$. Розіб'ємо цю множину на такі підмножини X_i , $i = \overline{0, 5}$:

- $X_0 = \{0\}$
- $X_1 = \{1\}$
- $X_2 = \{2\}$
- $X_3 = \{3\}$
- $X_4 = \{4, 5\}$
- $X_5 = \{6, 7, \dots\}$

Обчислимо ймовірності $p_i = \mathbb{P}(\xi \in X_i/H_0)$ та n_i - кількість значень реалізації вибірки, що потрапили в X_i .

X_i	$\{0\}$	$\{1\}$	$\{2\}$	$\{3\}$	$\{4, 5\}$	$\{6, 7, \dots\}$
p_i	0.2463	0.1856	0.1399	0.1055	0.1394	0.1833
$n \cdot p_i$	24.63	18.56	13.99	10.55	13.94	18.33
n_i	29	22	9	11	10	19

Бачимо, що $\sum_{i=0}^5 p_i = 1$, $r = 6$, і виконується умова $\forall i : np_i \geq 10$. Обчислимо значення статистики:

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

$$\begin{aligned} \eta_{\text{знач.}} &= \frac{(29 - 24.63)^2}{24.63} + \frac{(22 - 18.56)^2}{18.56} + \frac{(9 - 13.99)^2}{13.99} + \\ &+ \frac{(11 - 10.55)^2}{10.55} + \frac{(10 - 13.94)^2}{13.94} + \frac{(19 - 18.33)^2}{18.33} \approx 4.35007 \end{aligned}$$

$r - s - 1 = 6 - 1 - 1 = 4$, $\alpha = 0.05$, тому за таблицею розподілу Пірсона знайдемо значення $t_{0.05,4} = 9.5$. Бачимо, що $\mu_{\text{знач.}} < t_{0.05,4}$. Робимо висновок, що на рівні значущості 0.05 дані не суперечать висунутій гіпотезі про те, що генеральна сукупність розподілена за законом Паскаля ($\xi \sim Pas(3.06)$).