МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
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 курсу "Математична статистика"
Варіант – 126 (26)

Виконав:
студент КА-96
Терещенко Денис, КА-96

КИЇВ - 2021

Завдання варіанту.

За вихідні розподіли при отриманні вибірок брались розподіли шести типів:

- гауссівський,
- рівномірний,
- експоненціальний зі зсувом — щільність розподілу має вигляд:

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

- біноміальний,
- Пуассона,
- геометричний у формі $p_{\text{Geom}(\alpha)}(k) = \frac{\alpha^k}{(\alpha+1)^{k+1}}, k \in \mathbb{N}$

1) Проведіть первинний аналіз вибірки. Це включає статистичний ряд (для неперервних розподілів — інтервальний), емпіричну функцію розподілу (для неперервних розподілів — інтервальну), її графік, полігон частот (для дискретних розподілів), гістограму (для неперервних розподілів), box-plot.
2) Знайдіть вибіркове середнє, вибіркову дисперсію, виправлену вибіркову дисперсію, вибіркову медіану, вибіркову моду, вибіркові коефіцієнти асиметрії та ексцесу.
3) Обґрунтуйте та висуньте (нову) гіпотезу про розподіл генеральної сукупності.
4) Методом моментів та методом максимальної вірогідності знайдіть оцінки параметрів розподілу.
5) Для кожного параметра кращу з цих двох оцінок перевірте на (асимптотичну) незміщеність, консистентність та ефективність.
6) Побудуйте довірчі інтервали надійністю 0.95 для параметрів розподілу.
7) Нарешті, перевірте висунуту гіпотезу про розподіл генеральної сукупності за допомогою критерію χ^2 . Якщо гіпотеза суперечить вибірковим даним, перейдіть до п. 3.
8) Висновок.

1. Первинний аналіз вибірки.

Згідно з варіантом 126, маємо таку вибірку з $n = 100$ елементів:

4 3 5 4 5 4 0 2 4 4 5 2 4 4 4 4 3 5 4 5 3 5 4 4 3 5 4 2 5 4 6 5 2 2 3 2 2 4 3 3 3 2 4 3 5 3 4 4 6 3 4 4 4 5 5 3 3 3 3 6 4 3 2 2 4 5 1 5 3 5 4 1 2 4 4 3 3 5 2 3 5 5 4 5 4 5 3 2 5 3 6 6 4 4 3 2 4 4 4 1

Відсортована вибірка (варіаційний ряд) має вигляд (позначимо $x_i, i = \overline{1, 100}$):

0 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6

Бачимо, що розподіл г.с. є дискретним. Тоді нехай $x_1^*, \ldots, x_m^*$ – елементи вибірки, впорядковані за зростанням, причому кожне значення вказується лише один раз, n_k – число разів появи x_k^* в реалізації вибірки. n_k називається *частотою* появи x_k^*.

Зауважимо, що $n_1 + \cdots + n_m = n$.

Сума частот елементів $\sum_{i=1}^{k} x_i^*$ називається *кумулятивною частотою* n_k^*:

$$n_k^* = n_1 + ... + n_k$$

Величина $\nu_k = \dfrac{n_k}{n}$ називається *відносною частотою.*

Сума відносних частот елементів $\sum_{i=1}^{k} \nu_i$ називається *кумулятивною відносною частотою* елементу x_k^*. Отримаємо **статистичний ряд:**

Значення (x_k^*)	Частоти (n_k)	Кумулятивні частоти (n_k^*)	Відносні частоти (ν_k)	Кумулятивні відносні частоти (ν_k^*)
0	1	1	0.01	0.01
1	3	4	0.03	0.04
2	14	18	0.14	0.18
3	23	41	0.23	0.41
4	33	74	0.33	0.74
5	21	95	0.21	0.95
6	5	100	0.05	1.00

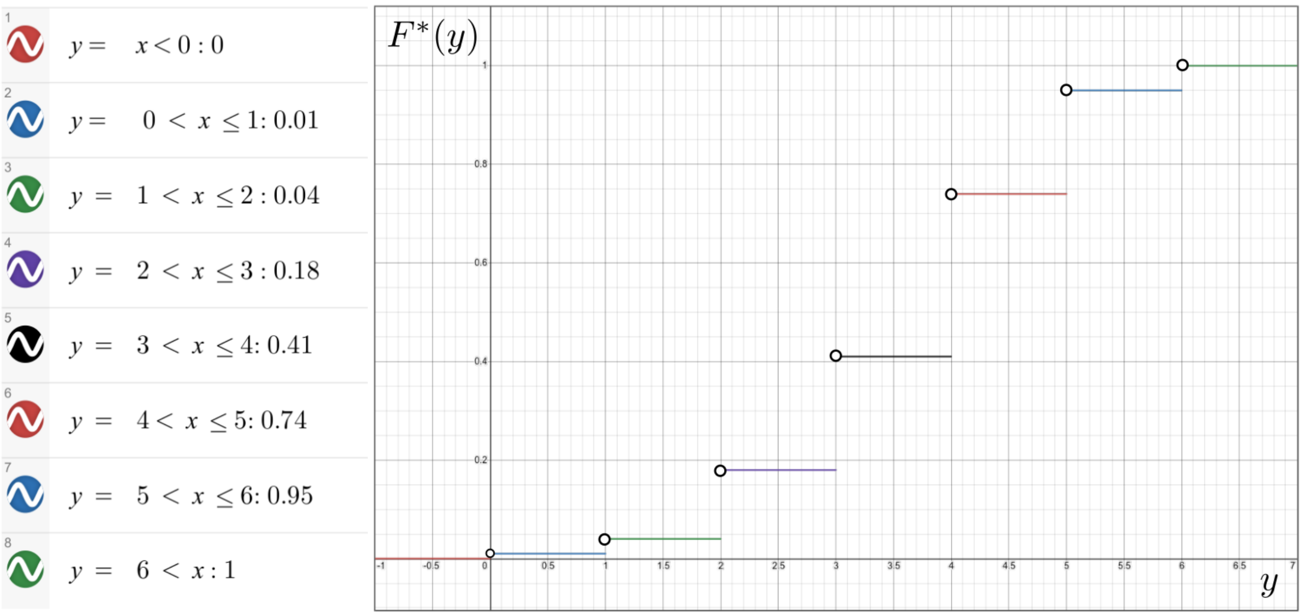
Емпірична функція розподілу.

Емпіричною функцією розподілу, побудованою за вибіркою $\xi_1, \ldots, \xi_n$ об'єму n, називається випадкова функція $F_n^* : \mathbb{R} \times \Sigma \to [0, 1]$ при кожному $y \in \mathbb{R}$ рівна:

$$F_n^*(y) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{\xi_i < y\}$$

Побудуємо для нашої вибірки:

$$F_{100}^*(y) = \begin{cases} 0, & y \leq 0; \\ 0.01, & 0 < y \leq 1; \\ 0.04, & 1 < y \leq 2; \\ 0.18, & 2 < y \leq 3; \\ 0.41, & 3 < y \leq 4; \\ 0.74, & 4 < y \leq 5; \\ 0.95, & 5 < y \leq 6; \\ 1, & y > 6. \end{cases}$$

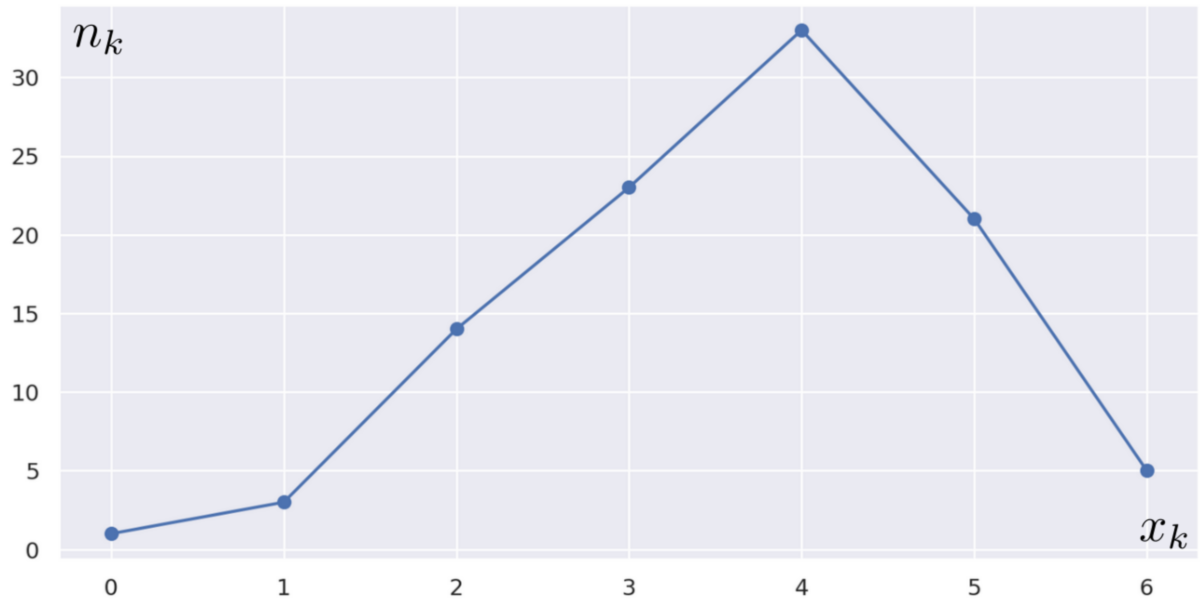


Полігон частот (count-plot).

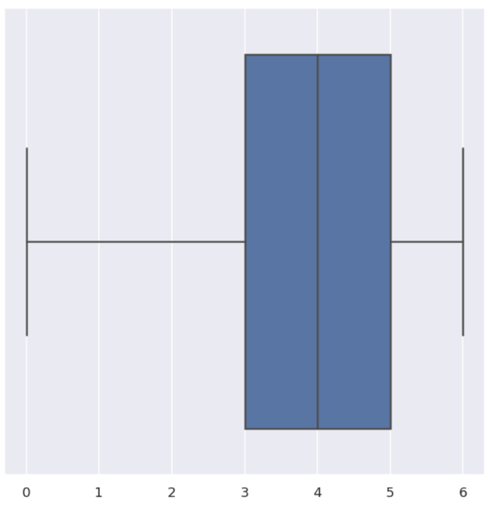
Якщо розподіл г.с. є дискретним, тоді полігон частот будується на основі статистичного розподілу г.с. наступним чином: будують систему координат таку, що на осі абсцис будуть відображатися елементи вибірки $x_1^*, \ldots, x_m^*$, а на осі ординат відповідні частоти.

Далі у вказаній системі координат будують точки $M_k(x_k^*, n_k), k = \overline{1, m}$, які з'єднують між собою у ламану $M_1 M_2 \ldots M_m$.

Полігон частот для заданої вибірки x



Boxplot.



The main component is the box, whiskers and outliers(number of individual points).

Components description

- Box - interquartile spread - [Q1, Q3].
- Vertical line inside the box - the median.
- Whiskers - [Q1 - 1.5 IQR , Q3 + 1.5 IQR].
- outliers - individual points.

2. Дескриптивні міри.

Знайдіть вибіркове середнє, вибіркову дисперсію, виправлену вибіркову дисперсію, вибіркову медіану, вибіркову моду, вибіркові коефіцієнти асиметрії та ексцесу.

- Вибіркове середнє: $\overline{x} = \frac{1}{100}\sum\limits_{i=1}^{100} x_i = \frac{367}{100} = 3.67$
- Вибіркова дисперсія: $\mathbb{D}_\xi^{**} \approx \frac{152.11}{100} = 1.5211$
- Виправлена вибіркова дисперсія: $\mathbb{D}_\xi^{***} = \frac{n}{n-1}\mathbb{D}_\xi^{**} \approx 1.5211 * \frac{100}{99} \approx 1.5364$
- Вибіркова мода – очевидно, що найчастіше зустрічається $\mathbf{4} = M_o$
- Вибіркова медіана: $M_e = \dfrac{x_{50} + x_{51}}{2} = 4$
- Коефіцієнт ассиметрії: $A_s = \dfrac{\overline{\mu}_3}{\sigma^3} \approx \frac{-0.64817}{1.87601} \approx -0.3455$
- Коефіцієнт ексцесу: $E_k = \dfrac{\overline{\mu}_4}{\sigma^4} - 3 \approx 2.8562 - 3 = -0.14373$

Проміжний висновок: розподіл скошено вліво (left skewed). Про це також свідчать положення вибіркових середнього та медіани. Усі значення у вибірці додатні, тож скористаємося коефіцієнтом варіації:

$$C_V = \frac{\sqrt{\mathbb{D}_\xi^{***}}}{\overline{x}} \cdot 100\% = 33\%$$

Це свідчить про суттєве розсієння ознаки по відношенню до середнього показника. Також маємо від'ємний коефіцієнт ексцесу, тож розподіл більш “пласковершинний” ніж нормальний.

3. Гіпотеза.

Обґрунтуйте та висуньте гіпотезу про розподіл генеральної сукупності.

Як було зауважено раніше, вибірка отримана з г.с. з дискретним розподілом. Розглянемо можливі варіанти:

○ Poisson distribution. Має додатню ассиметрію ($Sk = \lambda^{-\frac{1}{2}}$) та додатній коефіцієнт ексцесу $\lambda^{-1}, \lambda > 0$. Крім того, теоретичне значення середнього та дисперсії збігаються для розподілу Пуассона. Для нашої вибірки такі твердження будуть вкрай невірними. Таким чином, з великою ймовірністю вибірка прийшла не з розподілу Пуассона.

○ Geometric distribution. Для геометричного розподілу за законом:

$$p_{\mathrm{Geom}(\alpha)}(k) = \frac{\alpha^k}{(\alpha+1)^{k+1}}, k \in \mathbb{N}$$

характерні додатні коефіцієнти ассиметрії та ексцесу: $As = \frac{2-p}{\sqrt{1-p}}$ $Ek = 6 + \frac{p^2}{1-p}$. Також, теоретично мода геометричного розподілу дорівнює 1. Ці твердження не відповідають чисельним характеристикам нашої вибірки, тому навряд генеральна сукупність має геометричний розподіл.

○ Найбільш логічним припущенням в данному випадку буде *Біноміальний розподіл.* За чисельними характеристиками: від'ємний коефіцієнт ассиметрії та ексцесу свідчить, що $q < p$. Для порівняння теоретичного розподілу із розподілом вибірки, припустимо, за методом моментів, що:

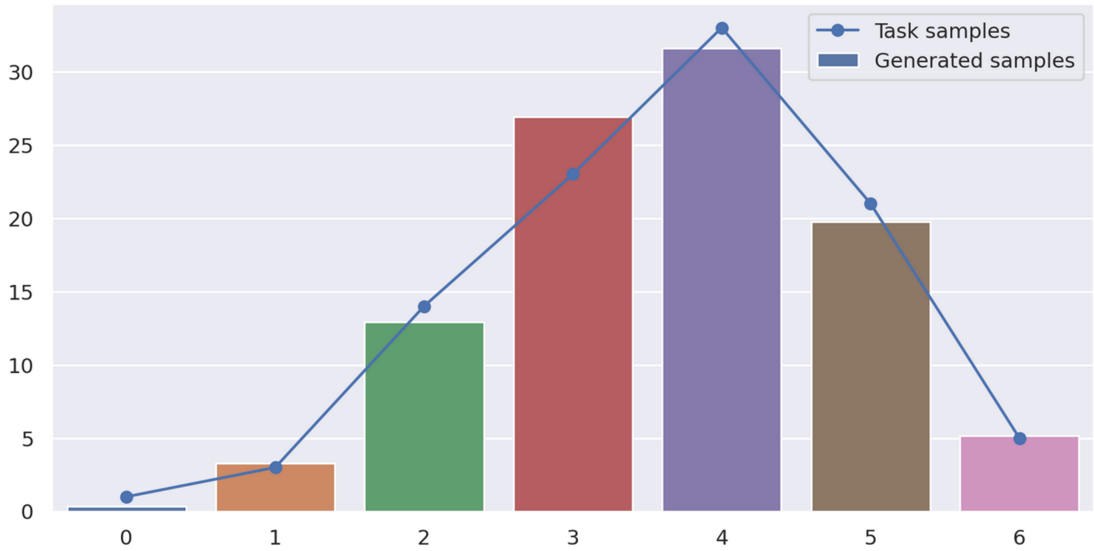
$$\xi \sim Bin(n,p) \qquad \begin{matrix} \mathbb{E}\xi = \overline{x} \\ \Downarrow \\ np = \overline{x} = 3.67 \end{matrix} \qquad \begin{matrix} \mathbb{D}_\xi = \mathbb{D}_\xi^{**} \\ \Downarrow \\ npq = 1.5211 \end{matrix} \implies \begin{cases} n = \lfloor \frac{a^2}{a-\sigma} \rfloor \approx 6 \\ p = 1 - \frac{\sigma}{a} \approx 0.59 \end{cases}$$

За нашим припущенням г.с.: $\xi \sim Bin(6, 0.61)$. Чому саме 0.61 – пояснимо далі. Навідь за такої грубої оцінки отримали "близькі" параметри:

$$As_\xi \approx -0.31 \qquad Ek_\xi \approx -0.29 \qquad Me_\xi = 4 \qquad Mo_\xi = \lceil 4.28 \rceil = 4$$

Змоделюємо вибірку обсягом 100 із біноміального розподілу та порівняємо полігони частот заданої вибірки та отриманої:

Згенеровані дані у порівнянні з заданою вибіркою:



Проміжний висновок: керуючись роздумами вище, можемо припустити, що г.с. має біноміальний розподіл. Пізніше більш корректно оцінемо параметри та перевіримо цю гіпотезу задопомогою критерію χ^2. Остаточно сформулюємо:

Гіпотеза #1

$$H_0 = \{\xi \sim \text{Bin}(n, p)\} \qquad H_1 = \{\xi \nsim \text{Bin}(n, p)\}$$

де ξ – вип. вел. Г.С., а параметри n, p – оцінемо у подальшому.

4. Оцінка параметрів.

Метод моментів.

Раніше виводили:

$$\xi \sim Bin(n,p) \qquad \begin{matrix} \mathbb{E}\xi = \overline{x} \\ \Downarrow \text{ м.м.} \\ np = \overline{x} = 3.67 \end{matrix} \qquad \begin{matrix} \mathbb{D}_\xi = \mathbb{D}_\xi^{**} \\ \Downarrow \text{ м.м.} \\ npq = 1.5211 \end{matrix} \implies \begin{cases} n = \lfloor \frac{a^2}{a-\sigma} \rfloor \approx 6 \\ p = 1 - \frac{\sigma}{a} \approx 0.59 \end{cases}$$

Але, т.я. ми зменшуємо n, округлюючи вниз, то точніше буде рахувати p від n:

$$np = \overline{x} \quad \implies \quad p = \frac{\overline{x}}{n} \approx 0.61$$

MLE.

Запишемо функцію вірогідності:

$$\mathcal{L}(\overrightarrow{x}_m, n, p) = \prod_{i=1}^{m} f(x_i) = \prod_{i=1}^{m} \left(\frac{n!}{x_i!\,(n-x_i)!} \right) p^{x_i} (1-p)^{n-x_i} =$$
$$= \left(\prod_{i=1}^{m} \left(\frac{n!}{x_i!\,(n-x_i)!} \right) \right) p^{\sum\limits_{i=1}^{m} x_i} (1-p)^{n - \sum\limits_{i=1}^{n} x_i}$$

Опустимо перший множник, так як він не залежить від p:

$$\ln \mathcal{L}(\overrightarrow{x}_m, n, p) = \sum_{i=1}^{n} x_i \ln(p) + \left(n - \sum_{i=1}^{n} x_i \right) \ln(1-p)$$

$$\frac{\partial \ln \mathcal{L}(\overrightarrow{x}_m, n, p)}{\partial p} = \frac{1}{p} \sum_{i=1}^{n} x_i + \frac{1}{1-p} \left(n - \sum_{i=1}^{n} x_i \right) = 0$$

$$(1-p^*) \sum_{i=1}^{n} x_i + p^* \left(n - \sum_{i=1}^{n} x_i \right) = 0 \implies \boxed{p^* = \frac{\sum\limits_{i=1}^{n} x_i}{n} = \frac{\overline{x}}{n}}$$

Маємо оцінку для p при відомому n. Маючи достатньо велике $m = 100$, виходячи з функції вірогідності, порівняно непоганою могла би бути оцінка:

$$n^* = \max_i x_i^* = 6$$

Скористаємося літературою, яка описує оцінку параметрів при невідомих n, p:

Estimation of Binomial Parameters when Both n and p are Unknown
A.DasGupta Herman Rubin
Purdue University February 8,2004

Один з найпростіших підходів, які розглядаються (у наших позначеннях):

$$\max_i X_i^* \xrightarrow[m\to\infty]{a.s.} n$$

Але, для застосування з прийнятною ймовірністю має бути: $m \geq 31500$. Слід зауважити, що у главі 2 (Nonexistence of Unbiased Estimates) представлено доведення відсутності незміщеної оцінки n при невідомих n, p.

Скористаємося оцінкою n, яку пропонує автор матеріалу:

$$n^* = \frac{x_{(k)}^{\alpha+1} \cdot s^{2\alpha}}{\overline{x}^{\alpha}(x_{(k)} - \overline{x})^{\alpha}}$$

де $x_{(k)}$ – максимум вибірки, $\alpha = \mathbb{P}\left\{x_{(k)} < m\right\}$. При підрахунку з $\alpha = 0.05$:

$$n^* \approx \lfloor 6.0039 \rfloor = \mathbf{6}$$

Проміжний висновок: таким чином, отримали вирази та чисельні оцінки для параметрів нашого гіпотетичного розподілу:

$$n^* \approx 6 \qquad p^* \approx 0,6116$$

Як вже було зазначено, оцінка n є зміщеною, тобто не є ефективною. Тому, будемо перевіряти властивості оцінки $p^* = \frac{\overline{x}}{n}$.

5. Властивості параметрів.

Незміщеність

Для кожного параметра кращу з цих двох оцінок перевірте на (асимптотичну) незміщеність, консистентність та ефективність. θ^* — **незміщена** оцінка параметру θ, якщо $\mathbb{E}\theta^* = \theta$.

$$\mathbb{E}p^* = \mathbb{E}\left(\frac{\overline{x_m}}{n}\right) = \mathbb{E}\left(\frac{\sum\limits_{i=1}^{m}\xi_i}{mn}\right) = \frac{nmp}{nm} = p \Longrightarrow \ \textit{незміщена}$$

Консистентність.

Означення. θ_n^* називається *консистентною* оцінкою параметра θ, якщо:

$$\theta_n^* \xrightarrow[n\to\infty]{\mathbb{P}} \theta \quad - \textit{слабка} \qquad\qquad \theta_n^* \xrightarrow[n\to\infty]{\text{м.н.}} \theta \quad - \textit{сильна}$$

Для параметру p^* за законом великих чисел маємо *консистентність*:

$$\lim_{m\to\infty} p_m^* = \lim_{m\to\infty} \frac{\frac{\xi_1+\cdots+\xi_m}{m}}{n} = \frac{1}{n}\lim_{m\to\infty}\frac{\xi_1+\cdots+\xi_m}{m} = |\text{м.н., ПЗВЧ}| = \frac{1}{n}\mathbb{E}\xi = p$$

Ефективність.

Оцінка для n зміщена, тобто неефективна.

Якщо $\exists C_{m,p} : C_{m,p}\cdot(p^*-p) = \frac{\partial \ln \mathcal{L}(\overline{x},p)}{\partial p}$, то оцінка p^* буде ефективною:

$$\frac{\partial \ln \mathcal{L}(\overline{x},p)}{\partial p} = \frac{\sum\limits_{i=1}^{m} x_i}{p} - \frac{mn - \sum\limits_{i=1}^{m} x_i}{1-p} = \frac{nmp^*}{p} - \frac{mn - mnp^*}{1-p} =$$
$$= mn\left(\frac{p^* - pp^* - p + pp^*}{p(1-p)}\right) = \frac{mn}{p(1-p)}(p^*-p)$$

Отже, $p^* = \dfrac{\overline{x}}{n}$ – *ефективна оцінка* параметра p т.я. $\exists C_{m,p} = \dfrac{mn}{p(1-p)}$.

6. Довірчі інтервали.

Побудуємо довірчий інтервал надійністю 0.95 для $p^* = \frac{\overline{x}}{n}$, де n – параметр біноміального розподілу, m – кількість елеметів заданої вибірки $\overrightarrow{x}_m = (x_1 \ \dots \ x_m)$.

При великих $m > 100$ за ЦГТ: $p^* \approx N(\mathbb{E}p^*, \mathbb{D}p^*) \Longrightarrow \dfrac{p^* - \mathbb{E}p^*}{\sqrt{\mathbb{D}p^*}} \approx N(0,1)$.

$$\mathbb{E}p^* = p \text{ (незміщеність)} \qquad \mathbb{D}p^* = \mathbb{D}\frac{\xi_1 + \dots \xi_m}{nm} = \frac{p - p^2}{m}$$

$$\frac{(p^* - p)\sqrt{m}}{\sqrt{p - p^2}} = \frac{p^* - p}{\sqrt{\frac{p-p^2}{m}}} \approx N(0,1)$$

$$\mathbb{P}\left\{-t_\gamma < \frac{(p^* - p)\sqrt{m}}{\sqrt{p - p^2}} < t_\gamma\right\} \approx \gamma \qquad \Phi(t_\gamma) = \gamma/2 \Rightarrow t_\gamma = 1.96$$

$$\frac{(p^* - p)^2 \cdot m}{p - p^2} < t_\gamma^2 \quad \Longleftrightarrow \quad mp^2 - 2mpp^* + m(p^*)^2 < t_\gamma^2 \cdot (p - p^2)$$

$$\left(1 + \frac{t_\gamma^2}{m}\right) p^2 - \left(2p^* + \frac{t_\gamma^2}{m}\right) p + (p^*)^2 < 0$$

$$D = 4(p^*)^2 + 4\frac{p^* t_\gamma^2}{m} + \frac{t_\gamma^4}{m^2} - 4(p^*)^2 - 4\frac{(p^*)^2 t_\gamma^2}{m} = t_\gamma^2 \left(\frac{4p^*}{m} + \frac{t_\gamma^2}{m} - \frac{4(p^*)^2}{m}\right)$$

$$p_{1,2} = \frac{2p^* + \frac{t_\gamma^2}{m} \pm t_\gamma \sqrt{\frac{t_\gamma^2}{m} - \frac{4p^*(1-p^*)}{m}}}{2\left(1 + \frac{t_\gamma^2}{m}\right)}$$

m – велике, тому можемо знехтувати $\frac{t_\gamma^2}{m}, \frac{t_\gamma^2}{m^2}, \frac{2t_\gamma^2}{m}$:

$$p_{1,2} \approx \frac{2p^* \pm t_\gamma \sqrt{\frac{4p^*(1-p^*)}{m}}}{2} = p^* \pm t_\gamma \sqrt{\frac{p^*(1-p^*)}{m}} = 0.6116 \pm 0.0956$$

Отже, з довірчою ймовірністю $\gamma = 0.95$ параметр $p \in (0.5160, 0.7071)$.

7. Перевірка гіпотези.

Нарешті, перевірте висунуту гіпотезу про розподіл генеральної сукупності за допомогою критерію χ^2 . Виберемо рівень значущості $\alpha = 0.05$.

$$\text{Гіпотеза:} \quad \begin{array}{c} H_0 = \{\text{Г.С. має біноміальний розподіл } Bin(6, 0.6116)\} \\ H_1 = \{\text{Г.С. має НЕ біноміальний розподіл } Bin(6, 0.6116)\} \end{array}$$

$$\text{Проміжки:} \quad \Delta_0 = \{0\}, \Delta_1 = \{1\}, \Delta_2 = \{2\}, \ldots, \Delta_6 = \{6\}$$

Для біноміального розподілу розрахуємо ймовірності попадання у Δ_k (PMF):

$$PMF: \quad \binom{n}{k} p^k q^{n-k} = P_k$$

де p, n – параметри розподілу, m – кількість елементів вибірки, P_k – ймовірність потрапляння у Δ_k.

$C(6,0) = 1.0$	$p^k \cdot q^{n-k} \approx 0.00343$	$P_k = 0.00343$
$C(6,1) = 6.0$	$p^k \cdot q^{n-k} \approx 0.00541$	$P_k = 0.03244$
$C(6,2) = 15.0$	$p^k \cdot q^{n-k} \approx 0.00851$	$P_k = 0.12769$
$C(6,3) = 20.0$	$p^k \cdot q^{n-k} \approx 0.0134$	$P_k = 0.26808$
$C(6,4) = 15.0$	$p^k \cdot q^{n-k} \approx 0.02111$	$P_k = 0.31661$
$C(6,5) = 6.0$	$p^k \cdot q^{n-k} \approx 0.03324$	$P_k = 0.19942$
$C(6,6) = 1.0$	$p^k \cdot q^{n-k} \approx 0.05234$	$P_k = 0.05234$

	Δ_0	Δ_1	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6
n_k	1	3	14	23	33	21	5
$m \cdot P_k$	0.343	3.244	12.769	26.808	31.661	19.942	5.234

Для більшості значень виконується $mP_k > 5$. Обчислимо $\chi^2(m)$:

$$\chi^2(100) \approx 1.2585 + 0.0184 + 0.1187 + 0.5409 + 0.0566 + 0.0561 + 0.0105 \approx 2.0596$$

За теоремою Пірсона про критерій χ^2:

$$\chi^2(100) \approx 2.0596 < 12.6 = \chi^2_{6,0.05}$$

Для порівняння, з іншим рівнем значущості $\alpha = 0.95$:

$$\chi^2(100) \approx 2.0596 < 2.17 = \chi^2_{6,0.95}$$

Таким чином, на вибраному рівні значущості $\alpha = 0.05$ гіпотеза H_0 не суперечить вибірковим даним.

8. Висновок.

В ході виконання розрахункової роботи було проведено опрацювання, дослідження та аналіз заданої вибірки зі 100 елеметів. Результатом ранжування даних є наведений варіаційний ряд. Генеральна сукупність з великою ймовірністю є дискретною випадковою величиною. Проведено первинний аналіз вибірки шляхом побудови статистичного ряду, емпіриної функції розподілу:

$$F_{100}^{*}(y) = \begin{cases} 0, & y \leq 0; \\ 0.01, & 0 < y \leq 1; \\ 0.04, & 1 < y \leq 2; \\ 0.18, & 2 < y \leq 3; \\ 0.41, & 3 < y \leq 4; \\ 0.74, & 4 < y \leq 5; \\ 0.95, & 5 < y \leq 6; \\ 1, & y > 6. \end{cases}$$

... та її графіка. За полігоном частот можна було передбачити від'ємний коефіцієнт ассиметрії та відносно великий розкид даних (що показує коефіцієнт варіації = 33%). Це свідчить про суттєве розсієння ознаки по відношенню до середнього показника. Також маємо від'ємний коефіцієнт ексцесу, тож розподіл більш "пласковершинний" ніж нормальний. Було розглянуто можливість 3х різних розподілів: пуассонівського, геометричного та біноміального. Перші 2 було відкинуто за несумісність з вибірковими даними (фактично, чисельні характеристики та форма графіків полігону частот та емпіричної функції розподілу є нехарактерними для даних розподілів). Для біноміального розподілу легко побачити схожість згенерованих даних з вибірковими.

Таким чином, була висунута гіпотеза про те, що вибірка прийшла з *біноміального розподілу.* Досить точною виявилася оцінка параметрів n, p методом моментів:

$$p^* = \frac{\overline{x}}{n} \qquad n = \lfloor \frac{a^2}{a - \sigma} \rfloor$$

Оцінка p^* – незміщена, консистентна, ефективна. Оцінка n – зміщена. Далі побудували довірчі інтервали для p^*: з довірчою ймовірністю $\gamma = 0.95$ параметр $p \in (0.5160, 0.7071)$.

Внаслідок перевірки гіпотези $H_0 = \{\text{Г.С.} = Bin(6, 0.6116)\}$ задопомогою критерію χ^2, було встановлено, що на вибраному рівні значущості $\alpha = 0.05$ гіпотеза H_0 не суперечить вибірковим даним.