

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА
ФАКУЛЬТЕТ ПРИКЛАДНОЇ МАТЕМАТИКИ ТА ІНФОРМАТИКИ

Кафедра дискретного аналізу та інтелектуальних систем

Індивідуальне завдання №2
з курсу «Теорія ймовірності та математична статистика»

Виконав:

Студент групи ПМІ-23

Лук'янчук Денис

Викладач:

Доц. Пелюшкевич Ольга

Володимирівна

Постановка задачі

Метою даного дослідження є перевірка відповідності вибірки даних, зчитаних із файлу, теоретичному розподілу Пуассона. Необхідно визначити, чи можна стверджувати, що дані розподілені за законом Пуассона з параметром λ , оцінити цей параметр із вибірки або використати заданий вручну, а також провести статистичний аналіз за критерієм χ^2 із рівнем значущості α . У процесі аналізу передбачається об'єднання класів із частотами меншими за 5 для забезпечення коректності тесту, візуалізація результатів та оцінка внеску кожного класу в загальну статистику.

Теорія

Для перевірки гіпотези за критерієм Пірсона виконаємо такі кроки:

1. Складемо статистичні дані для чисової вибірки у вигляді інтервального статистичного розподілу.
2. Визначимо теоретичні ймовірності появи значень вибірки відповідно до закону розподілу Пуассона.
3. Обчислимо емпіричне значення критерію узгодженості Пірсона за формулою:

$$\chi_{\text{emp}}^2 = \sum_{i=1}^m \frac{(n_i - np_i)^2}{np_i}$$

4. За заданим рівнем значущості визначимо кількість ступенів свободи $k = m - s - 1$, де m — кількість інтервалів, а s — кількість параметрів, які обчислюються за таблицею критичних значень розподілу χ^2 .
5. Знайдемо емпіричне та критичне значення. Якщо емпіричне значення більше за критичне, то гіпотезу відхиляємо. Якщо емпіричне значення менше за критичне, гіпотезу приймаємо.

Закони розподілу

1. Біноміальний закон розподілу:

$$p_i = P(\xi = i) = C_N^i p^i (1-p)^{N-i} \quad p = \frac{\bar{x}}{N}$$

2. Закон розподілу Пуассона:

$$P(\xi = i) = e^{-\lambda} \frac{\lambda^i}{i!} \quad \lambda > 0$$

3. Рівномірний закон розподілу:

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases}$$

$$a = \bar{x} - \sqrt{3}s, \quad b = \bar{x} + \sqrt{3}s$$

4. Показниковий закон розподілу:

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x \geq 0 \end{cases} \quad \lambda = \frac{1}{\bar{x}}$$

5. Нормальний закон розподілу:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-a)^2}{2\sigma^2}} dt \quad a = \bar{x}, \quad \sigma = s$$

Хід роботи

1. Вибір вхідних даних

- Відповідно до номера варіанту в журналі, обирається файл із вхідними даними sample14.txt. Дані читаються з текстового файлу для подальшого аналізу.

2. Постановка задачі

- Зчитати вибірку з файлу.
- Опрацювати статистичний матеріал, представивши дані у табличній та графічній формах.
- На основі графічного представлення сформулювати гіпотезу про закон розподілу генеральної сукупності.
- Надати користувачу можливість задавати параметри розподілу вручну або оцінити їх автоматично на основі вибірки.
- Перевірити сформульовану гіпотезу за критерієм χ^2 для заданого рівня значущості α .

3. Аналіз вибірки

- Визначити розмір вибірки n мінімальний та максимальний елементи.
- Користувач задає рівень значущості α та обирає спосіб визначення параметра розподілу (вручну або автоматично).

4. Обробка даних

- Зчитування вибірки:
 - Дані з файлу читаються у список чисел.
- Обчислення базових параметрів:
 - Розрахунок середнього значення вибірки для автоматичної оцінки параметра розподілу (наприклад, `lambda`) для Пуассона).
 - Якщо обрано "вручну", користувач вводить значення параметра; інакше використовується автоматична оцінка.

Вибір параметра λ :

Автоматично (середнє вибірки)
 Вручну

Введіть λ (якщо вручну):

Рівень значущості α :

Обробити дані

3. Побудова частотної таблиці:

- Визначення унікальних значень вибірки та їх частот (спостережувані частоти m_i).
- Обчислення теоретичних ймовірностей p_i для обраного розподілу.
- Розрахунок очікуваних частот $n p_i$.
- Коригування категорій: об'єднання тих, де m_i або $n p_i$ менші за 5.

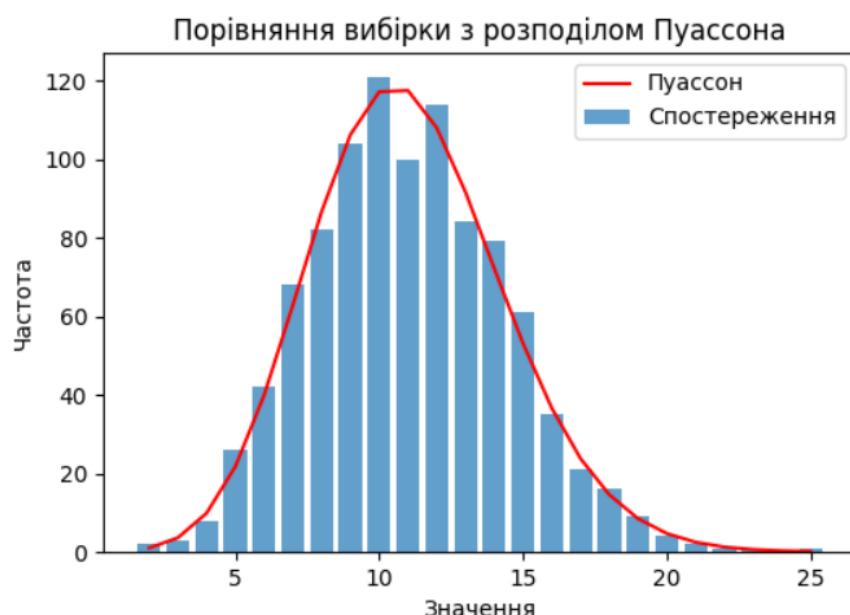
Частотна таблиця

x_i	m_i	p_i	n p_i
4	13	0.0146	14.37
5	26	0.0220	21.69
6	42	0.0405	39.88
7	68	0.0639	62.84
8	82	0.0881	86.65
9	104	0.1079	106.20
10	121	0.1190	117.14
11	100	0.1194	117.46
12	114	0.1097	107.97
13	84	0.0931	91.61
14	79	0.0734	72.18
15	61	0.0539	53.08
16	35	0.0372	36.59
17	21	0.0241	23.74
18	16	0.0148	14.55

4. Графічне представлення:

- Побудова гістограми спостережуваних частот.
- Накладання кривої теоретичного розподілу з обраним параметром.

Гістограма



5. Формулювання гіпотези:

- На основі графіку оцінюється схожість вибірки з теоретичним розподілом.
- Формулюється нульова гіпотеза H_0 : дані відповідають обраному розподілу.

6. Перевірка гіпотези за критерієм χ^2 :

- Обчислення емпіричного значення χ^2 емпіричне за скоригованими частотами.
- Визначення ступенів свободи df .
- Отримання критичного значення χ^2 критичне для заданого $alpha$ та df .
- Порівняння χ^2 емпіричне з χ^2 критичне для висновку про прийняття або відхилення H_0 .

Розмір вибірки: 984
Оцінене λ (за вибіркою): 11.030
Використовується $\lambda = 11.030$
Рівень значущості α : 0.05

На основі графіку: вибірка нагадує розподіл Пуассона
 H_0 : Дані відповідають розподілу Пуассона з $\lambda = 11.030$

χ^2 (Емпіричне значення) = 7.90
 df (Ступені свободи): 15
 χ^2 (Критичне значення) ($\alpha = 0.05$; $df = 15$): 24.9958
Висновок: χ^2 (Емп) < χ^2 (Кр) – гіпотезу приймаємо

Висновок: У ході виконання завдання було проаналізовано вибірку з файлу, відповідного номеру варіанту в журналі. Дані успішно зчитано, опрацьовано та представлено у вигляді частотної таблиці та графіку, що дозволило візуально оцінити їх розподіл. На основі графічного аналізу сформульовано гіпотезу про відповідність вибірки певному закону розподілу (наприклад, Пуассона), яка була перевірена за критерієм χ^2 із заданим користувачем рівнем значущості α . Програма забезпечила гнучкість у виборі параметрів розподілу — користувач міг задавати їх вручну або використовувати автоматичну оцінку на основі вибірки. Результати статистичного тесту показали, чи є підстави прийняти нульову гіпотезу, що підтверджує відповідність даних передбаченому розподілу. Таким чином, поставлені задачі виконано: проведено повний аналіз вибірки, від табличного представлення до статистичного висновку, із застосуванням сучасних інструментів обробки даних.