

УДК 532.696.1

Н.Б. Долішня, Н.М. Піндус, С.А. Чеховський

Івано-Франківський національний технічний університет нафти і газу, Івано-Франківськ

ЗАСТОСУВАННЯ ПАРАМЕТРИЧНИХ ПІДХОДІВ ДЛЯ ПЕРЕВІРКИ ЗАКОНУ РОЗПОДІЛУ СТАТИСТИЧНИХ ДАНИХ В УМОВАХ ОБМЕЖЕНОГО ОБ'ЄМУ ВИБІРОК

Розглядаються методи перевірки закону розподілу малих вибірок за експериментальними даними одержаними в результаті повірки лічильників газу в умовах «Івано-Франківськгаз». Досліджено доцільність застосування графічного, аналітичного та інформаційного підходів перевірки виду кривої закону розподілу.

Ключові слова: мала вибірка, закон розподілу, графік Пірсона, ентропійний коефіцієнт.

Вступ

Постановка проблеми та аналіз останніх публікацій. Аналізуючи питання ідентифікації закону розподілу при опрацюванні результатів експериментальних досліджень у нафтогазовій промисловості, що, зокрема, стосується витрати газу, зазначимо некоректність застосування методів математичної статистики класичної теорії похибок [1 – 5]. Тому доцільною є розробка відповідного алгоритму опрацювання результатів експериментальних досліджень, враховуючи закони розподілу основних фізичних величин в даному опосередкованому вимірюванні витрати газу.

Актуальність цієї проблеми викликана труднощами не лише в процесі обробки статистичних даних, а й на шляху проектування та створення нової зразкової техніки або ж вдосконалення уже існуючих на сьогоднішній день засобів вимірювання. На практиці часто можливість ідентифікації форми розподілу експериментальних даних обмежена недостатньою кількістю елементів вибірки, адже при малому об'ємі вибірки особливості розподілу можуть виявитися «замаскованими» випадковістю самої ж вибірки [6].

Формулювання мети статті. Використовуючи наявні результати повірки лічильників різного типу та об'єму в умовах «Івано-Франківськгаз», постало завдання перевірки форми закону розподілу проміжних даних повірки на основі параметричних методів, а також визначення ефективного методу перевірки приналежності закону до певної групи розподілів.

Виклад основного матеріалу

На сьогоднішній день в математичній статистиці існує багато різноманітних критеріїв, що дають відповідь на питання ідентифікації закону розподілу отриманих даних, проте зупинимось на графічному, аналітичному на інформаційному підході, в основу яких закладена параметрична оцінка вибірки. У [2] згадується про деякі підходи, що застосовуються для підбору емпіричного розподілу ймовірностей при малих об'ємах вибірок. Відомо [6], що за допо-

могою графіка Пірсона, розрахувавши показники форми розподілу, а саме: коефіцієнти асиметрії β_1 та ексцесу β_2 , можна визначити тип закону розподілу даних. Перевага надається методу Пірсона в порівнянні з методами Шарльє і Крамера, завдяки тому, що охоплює практично всі відомі види статистичних розподілів. Коефіцієнти β_1 та β_2 виражаються через третій та четвертий центральний момент відповідно:

$$\beta_1 = \frac{\mu_3}{\sigma^3} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3 \left/ \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^{3/2} \right. ; \quad (1)$$

$$\beta_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4 \left/ \left[\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^2 \right. - 3, \quad (2)$$

де μ_3 – третій центральний момент; μ_4 – третій центральний момент; σ – середньоквадратичне відхилення (СКВ); x_i – i -й результат спостереження; \bar{x} – середнє арифметичне вибірки; n – об'єм вибірки (кількість елементів).

На основі отриманих статистичних даних повірки для кожного з 36 окремих файлів, які в собі містять результати повірок (у кожному з файлів містилося від 2 до 9 вибірок в залежності від типу та об'єму лічильника, що підлягав метрологічній повірці) розраховувались коефіцієнти асиметрії та ексцесу з метою «перенесення» отриманих значень на графік Пірсона та встановлення можливого закону розподілу даних. Проте на даному етапі дослідження виявлено, що розрахункові значення β_1 та β_2 є настільки малими ($\beta_1=(0;1)$ $\beta_2=(0;1)$), що робити висновок про вид закону розподілу на основі графіка Пірсона не можливо, адже вони знаходяться в критичній області.

Наступним методом перевірки виду функції закону розподілу є побудова гістограми. Як вже згадувалося вище, головним фактором, який ускладнює задачу ідентифікації форми кривої розподілу малих вибірок є сама випадковість появи різних значень випадкової величини. Звісно, уникнути цієї проблеми можна за умови збільшення об'єму вибірки, та іноді, як і в нашому випадку, це не є можливим за самою

суттю експерименту. Оперуючи невеликою кількістю елементів вибірки, необхідно перш за все спирається на максимальне використання апріорної інформації про вид розподілу, а саме береться до уваги дві апріорних аксіом: плавність та симетричність кривої розподілу [6], що видно із рисунку гістограми (рис. 1).

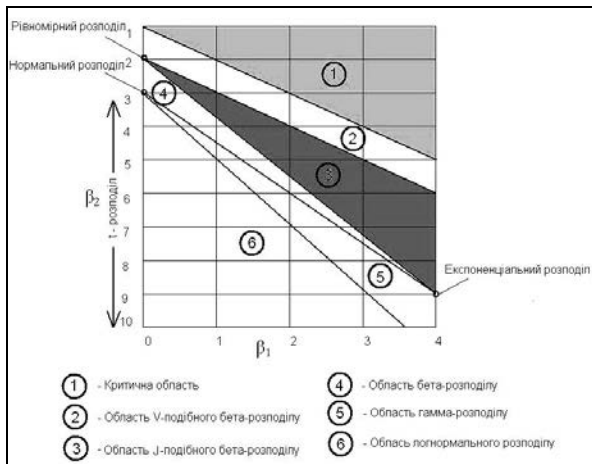


Рис. 1. Графік Пірсона для ідентифікації законів розподілу в площині $(\beta_1; \beta_2)$

Перейдемо до самої процедури побудови гістограми. Для найменшого спотворення кривої в області центру розподілу число стовпців рекомендується вибирати непарним, а центральний стовпець розміщувати симетрично відносно прийнятого центру розподілу, згідно з однією із вищезгаданих аксіом. Варто зауважити, що оптимальне число інтервалів вибраних для побудови гістограми сильніше залежить від форми розподілу аніж від об'єму вибірки, що видно із формули:

$$m = \frac{\varepsilon + 1.5}{6} n^{0.4} = \frac{1}{3} \sqrt[5]{n^2 / \chi^8}, \quad (3)$$

де m – кількість інтервалів; n – об'єм вибірки; ε – ексцес; χ – контрексцес.

З метою полегшення розрахунків і для графічного представлення дослідних даних створено програмне забезпечення (робоче вікно якої показано на рис. 2). Беручи до уваги рекомендації подані в [6], мінімальне і максимальне число інтервалів побудови гістограми вибиралося як:

$$m_{\min} = 0.55n^{0.4} \quad \text{та} \quad m_{\max} = 1.25n^{0.4}. \quad (4)$$



Рис. 2. Робоче вікно програми

Користувач може сам вибирати число інтервалів, тим самим збільшуючи або зменшуючи ширину кожного з інтервалів, враховуючи необхідність нерозривності інтервального ряду. Разом з тим, розраховані для попереднього методу точкові оцінки асиметрії та ексцесу, які мають асимптотично нормальний розподіл, використовувалися для перевірки належності вибірки до нормального закону розподілу на основі експрес-методу. Оскільки очікуваної результативності із використання графіка Пірсона не було отримано, то на даному етапі для кожної з вибірок формувався висновок про гостровершинність (плосковершинність) та правосторонню (лівосторонню) асиметрію кривої розподілу. Проте для середніх ($n = 50 \dots 60$) та малих ($n \leq 25$) вибірок розподіл оцінок асиметрії та ексцесу вже значно відхиляється від нормального, що різко знижує ефективність методу. Тому вибраного виду розподілу за формою побудованої гістограми та отриманого припущення на основі експрес-методу перевірки нормальності недостатньо для остаточного висновку.

Аналізуючи останні праці присвячені малим вибіркам [2], увагу привернула ідея застосування інформаційного підходу для визначення виду кривої закону розподілу, в основі якого лежить інформація, яку несе в собі досліджувана вибірка. Власне, саме інформаційний підхід дає змогу виявити «випадковість» набору даних. Критерієм приналежності кривої закону розподілу до того чи іншого сімейства розподілів служить ентропійний коефіцієнт k [6]:

$$k = \Delta_e / \sigma, \quad (5)$$

де Δ_e – ентропійне значення похибки.

У [6] можна знайти табличні значення коефіцієнта ентропії, за величиною якого вибірці приписують той чи інший закон розподілу.

Як видно із рис. 2 для кожної вибірки розраховувалися середнє значення, СКВ, коефіцієнти асиметрії, ексцесу, варіації та ентропії, так як саме ці параметри дають змогу робити висновок про закон розподілу, застосовуючи один із розглянутих вище методів.

На основі отриманих результатів робимо висновок, що графічний метод є нечутливим до вибірок малого об'єму або ж вплив додаткових факторів є настільки вагомим, що «маскує» реальні результати дослідження, оскільки на першому етапі перевірки закону розподілу зіткнулися з проблемою: за розрахунковими значеннями коефіцієнтів асиметрії та ексцесу неможливо зробити висновок про вид кривої розподілу, а також при ідентифікації закону $(\beta_1; \beta_2)$ виявлено, що кожна із дослідних вибірок знаходиться у критичній області, тобто віднести вибірку до конкретної групи розподілів не вдалося.

Тому наступним етапом було вирішено застосувати графічний підхід і охарактеризувати закон розподілу за допомогою гістограми, побудованої на основі наявних вибірок. Прослідковується певна закономірність і тенденція: для лічильників, які розраховані на малі витрати, притаманна переважання більшість плосковершинних законів розподілу з лі-

восторонньою асиметрією, а для більших витрат – з правосторонньою асиметрією. Разом з тим виявлено схильність до двохмодальності при малих об'ємах вибірок. Проте і тут однозначного висновку не отримано, оскільки було порушено кілька важливих принципів при побудові, що часто зумовлено малим обсягом та «випадковістю» самої вибірки. Важливою перевагою застосування графічного методу є те, що саме з рисунку гістограми можна зробити попереднє припущення про вид закону розподілу (віднести криву до певного класу розподілів) та про наявність можливих промахів.

Зважаючи на актуальність та широку застосовуваність інформаційного підходу, на завершальному етапі застосували саме його, що і дало змогу узагальнити результати отримані в ході експериментального дослідження доцільності застосування параметричних методів. Як і очікувалося, припущення винесені на другому етапі, підтверджуються результатами отриманими при розрахунку ентропійного коефіцієнту. Переважна більшість дослідних даних (близько 65%) підпорядковується нормальному закону розподілу, що підтверджується розрахунковими (за результатами розрахунку ентропійного коефіцієнта) та графічними (за гістограмою) даними, а також мають місце рівномірний, арксинусоїдальний, експоненціальний та t -розподіли (35%). Разом з тим не виявлено чіткої залежності між коефіцієнтом ентропії та об'ємом вибірки, а також підтверджений той факт, що більша кількість відтворених дослідів дає змогу точніше судити про вид закону розподілу отриманих даних загалом.

Висновки

Таким чином, на основі результатів перевірки лічильників різного типу та об'єму в умовах «Івано-Франківськгаз» із застосуванням аналітичного, графічного та інформаційного підходів було виявлено, що результативності та ефективності при визначенні виду кривої закону розподілу можна досягти при послідовному використанні кожного із зазначених методів, але прийти до однозначного висновку, застосувавши один із підходів, не вдається. Тому постає задача розробки оптимального методу визначення закону розподілу для вибірок малого об'єму із врахуванням результатів

отриманих вище. Підтверджено, що в умовах обмеженого об'єму вибірок важливим фактором є відтвореність експериментальних даних для можливості формування висновків про результат дослідження. Як і передбачалося, не всі вибірки підпорядковуються нормальному закону розподілу в силу багатьох різних за своєю фізичною суттю впливових факторів, що вимагає подальшого дослідження і розробок у напрямку опрацювання результатів вимірювання. Також подальшого дослідження потребують факт зміни закону розподілу на різних витратах лічильників, вплив зміни гідродинамічних процесів потоків та дослідження впливів цієї зміни на формування законів розподілу досліджуваних вибірок, що дасть змогу вдосконалити та оптимізувати алгоритм опрацювання результатів опосередкованого вимірювання витрати газу в умовах «Івано-Франківськгаз».

Список літератури

1. Алгоритм опрацювання даних для метрологічної атестації засобів вимірювань / А.К. Зюзько, М.Ю. Буриченко, Ю.В. Петрова, В.В. Німич // *Електроніка та системи управління*. – 2009. – №1(19). – С. 5-10.
2. Коваленко І.І. Оценки статистических характеристик вероятностных распределений малых выборок / И.И. Коваленко, Т.С. Гавриш // *Комп'ютерні технології. Наукові праці*. – Вип. 93, Т. 106. – С. 32-38.
3. Pearson E.S. Test for departure from normality: Comparison of powers / E.S. Pearson, R.B. D'Agostino, K.O. Bowman // *Biometrika*, 64, 1977. – P. 231-246.
4. Лемешко Б.Ю. Сравнительный анализ критериев проверки отклонения распределения от нормального закона / Б.Ю. Лемешко, С.Б. Лемешко // *Метрология*. – 2005. – № 2. – С. 3-23.
5. Долишня Н.Б. До питання застосування статистичних методів обробки експериментальних даних / Н.Б. Долишня, Н.М. Пиндус, С.А. Чеховський // *Приладобудування 2010: стан і перспективи: IX міжнар. наук.-техн. конф.*, 27 – 28 квітня 2010 р: зб. тез доповідей. – К.: НТУУ «КПІ». – 2010. – С. 105.
6. Новицький П.В., Зограф І.А. Оценка погрешностей результатов измерений. / П.В. Новицький, І.А. Зограф. – Л.: Энергоатомиздат, 2001. – 303 с.

Надійшла до редколегії 3.10.2010

Рецензент: д-р техн. наук, проф. І.П. Захаров, Харківський національний університет радіоелектроніки, Харків.

ПРИМЕНЕНИЕ ПАРАМЕТРИЧЕСКИХ ПОДХОДОВ ДЛЯ ПРОВЕРКИ ЗАКОНА РАСПРЕДЕЛЕНИЯ СТАТИСТИЧЕСКИХ ДАННЫХ В УСЛОВИЯХ ОГРАНИЧЕННОГО ОБЪЕМА ВЫБОРОК

Н.Б. Долишня, Н.М. Пиндус, С.А. Чеховський

Рассматриваются методы проверки закона распределения малых выборок по экспериментальным данным, полученным в результате проверки счетчиков газа в условиях «Ивано-Франковскгаз». Исследована целесообразность использования графического, аналитического и информационного подходов проверки вида кривой закона распределения.

Ключевые слова: малая выборка, закон распределения, графики Пирсона, энтропийный коэффициент.

THE APPLICATION OF PARAMETRIC APPROACHES FOR CHECKING THE DISTRIBUTION LAW OF THE STATISTICAL DATA LIMITED BY THE SAMPLES VOLUME

N.B. Dolishnia, N.M. Pindus, S.A. Chehovskiy

The methods of checking the distribution law of small test samples taken from the experimental data obtained by calibration the gas meters in Ivano-Frankivskgas were examined. The expediency of using the graphical, analytical and information approaches for checking the type of distribution curve was studied.

Keywords: small selection, distributing law, graphs of Pirson, coefficient of entropy.