УДК 528.11

Карпік Сергій, ст. магістратури факультету кібернетики; науковий керівник – д.ф.-м.н., професор Джунь Й. В. (Міжнародний економіко-гуманітарний університет імені академіка Степана Дем'янчука, Рівне)

ЗАКОН РОЗПОДІЛУ ПОХИБОК БАГАТОКРАТНИХ СПОСТЕРЕЖЕНЬ ВЕЛИКИХ ОБСЯГІВ ТА ОЦІНКА ЙОГО ПАРАМЕТРІВ

Анотація. *У статті досліджено важливу проблему прикладної математики – автоматизацію методів оцінки параметрів статистичних розподілів. Розкрито метод отримання ефективних параметрів розподілу Пірсона-Джеффріса. Показано значення та визначено математичні властивості цього універсального методу розподілу похибок спостережень великих обсягів. Для оцінки параметрів закону Пірсона-Джеффріса запропоновано сучасний програмний продукт в середовищі Borland C++ Builder 6. Запропонований продукт дає можливість за лічені секунди отримати оптимальні рішення.*
Ключові слова: *розподіл Пірсона-Джеффріса, оцінювання параметрів.*

Аннотация. *В статье исследовано важную проблему прикладной математики – автоматизацию методов оценки параметров статистических распределений. Раскрыто метод получения эффективных параметров распределения Пирсона-Джеффриса. Показано значение и определены математические свойства этого универсального распределения погрешностей наблюдений больших объемов. Для оценки параметров закона Пирсона-Джеффриса предложен современный программный продукт в среде Borland C + + Builder 6. Предложенный продукт дает возможность за считанные секунды получить оптимальные решения.*
Ключевые слова: *распределение Пирсона-Джеффриса, оценка параметров.*

Annotation*. The article is devoted to the solution of important problems in applied mathematics – automation methods estimate the parameters of statistical distributions. The theme of the research is to develop a method of obtaining effective distribution parameters of the Pearson-Jeffries. The paper shows the importance and Points mathematical properties of universal distribution of large amounts of observational errors. For estimates of the parameters of the law-Pearson Jeffries suggested the current software in the environment of Borland C + + Builder 6. Proposed product makes it possible for a few seconds to get the best solutions.*
Keywords: *The distribution of Pearson-Jefferies, estimates of the parameters.*

Основною особливістю сучасних наукових експериментів є великі, або надзвичайно великі обсяги вимірювальної інформації, яка їх забезпечує. Це обумовлено автоматизацією спостережень. Проте, математичне моделювання і обробка цих експериментів виконується на основі застарілих уявлень і застарілими методами. Отже, виникає проблема використання сучасних математичних процедур при аналізі великих обсягів даних.

Метою і завданням цього дослідження є ознайомлення широкого кола науковців з сучасними поглядами на математичну обробку інформації і з новим фундаментальним законом похибок, який використовується при обробці вибірок великих обсягів, а також з методами оцінки його параметрів.

Відомий англійський математик, геофізик і астроном Кембриджського університету професор сер Г. Джеффріс ще в 1937 р. прийшов до висновку, що класичний закон похибок Гауса виявляє свою повну теоретичну і практичну неспроможність за умови, якщо число багатократних спостережень n>500. Ця концепція Джеффріса з великими труднощами пробивала собі дорогу, оскільки закон Гауса, як фундаментальний принцип математичного моделювання, прекрасно зарекомендував себе на протязі більше ніж ста років його практичного використання. Більшість математиків вважали, що саме закон Гауса складає суть центральної граничної теореми теорії ймовірностей. Тому здавалося, що пропозиція Джеффріса знаходиться в дисонансі як з набутим досвідом математичної обробки даних, так і з самою теорією ймовірностей.

Оскільки сер Г. Джеффріс є дуже авторитетним вченим в НАН України в 70-х рр. минулого століття під керівництвом академіка Е. П. Федорова була здійснена фундаментальна перевірка згаданої вище концепції. Для такої перевірки були залучені спостереження найвищої якості починаючи з історичних рядів Ф. В. Бесселя [1] і закінчуючи найсучаснішими рядами космічних, астрономічних, геодезичних та інших спостережень [2–6]. На великий подив більшості вчених згадана перевірка підтвердила правильність концепції Джеффріса: ряди похибок з числом спостережень n>500 на 100 % мали негаусівський характер і підкорялись розподілу:

$$y(x)=\frac{\Gamma(m+1)}{\sqrt{2\pi(m-0{,}5)\cdot\Gamma(m+0{,}5)}}\cdot\frac{1}{\sigma_{VII}}\left[1+\frac{0{,}5m^2}{(m-0{,}5)^3}\left(\frac{x-a_{VII}}{\sigma_{VII}}\right)^2\right]^{-m}, \qquad (1)$$

де a_{VII}, σ_{VII}, m – параметри форми закону щільності імовірностей (1).

При чому, в якій би галузі науки не досліджувались похибки спостережень, в усіх випадках вони описувались магічною формою Джеффріса (1).

Цікавим є походження закону (1). Джеффріс називає цей закон розподілом Пірсона VII типу, який має недіагональну інформаційну матрицю, Джеффріс перетворив його до форми (1) яка, як і закон Гауса, має незалежні параметри.

Будучи незвичайно скромною людиною Джеффріс не дав особливої назви розподілу (1) який він створив, продовжуючи називати його розподілом Пірсона VII типу. Тому багато хто з дослідників, не знаючи цих тонкощів, ідентифікують класичну криву Пірсона VII типу і навіть узагальнений розподіл Коші [7] з формою (1), хоча цього робити не можна, оскільки це різні розподіли. Враховуючи сказане, розподіл (1) названо законом похибок Пірсона-Джеффріса, або просто законом Пірсона-Джеффріса, подібно тому, як у свій час був визначений закон Гауса.

Значення закону Пірсона-Джеффріса (1) для математичної обробки інформації, полягає в тому, що він має цінні математичні властивості:

1. Є узагальненням двох найбільш важливих в теорії помилок і аналізі даних розподілів: Гауса, у який формула перетворюється при $m=\infty$, і Стьюдента при значеннях m в (1) кратних 0,5.

2. Параметр m закону Пірсона-Джеффріса (1) і число ступенів свободи ν t-розподілу пов'язані відношенням [8]:

$$\nu = 2m - 1$$

3. Параметр m є чутливою мірою відхилення розподілу (1) від нормального закону. Гаусова форма розподілу (1) постулює значення m наперед відомим і рівним ∞. Насправді ж кожен ряд похибок характеризується своїм значенням m, яке досить далеке від ∞ і є його найважливішою характеристикою.

4. Найбільш цінною властивістю закону Пірсона-Джеффріса є діагональність його інформаційної матриці (як і у нормального розподілу). Зауважимо, багато розподілів, які пропонувались для заміни закону Гауса, наприклад, Lp- розподіл – такою властивістю не володіють.

5. Границі нерівності Рао-Крамера для закону Пірсона-Джеффріса є такими [9]:

$$\sigma_a^2 \geq \frac{\sigma_{VII}^2}{n} \cdot \frac{(m-0{,}5)^2(m+1)}{m^3}\ ;\ \sigma_\sigma^2 \geq \frac{\sigma_{VII}^2}{2n} \cdot \frac{(m+1)}{(m-0{,}5)}\ ; \qquad (2)$$

$$\sigma_m^2 \geq \left\{ n\left[\psi'(m-0{,}5) - \psi'(m) - \frac{m+1}{2m^2(m-0{,}5)} \right] \right\}^{-1} = \frac{\sigma_{1m}^2}{n}. \qquad (3)$$

де $\psi'(m)$ – тригама – функція.

Із формули (2) видно, що при $m=\infty$ (закон Гауса) σ_a^2 і σ_σ^2 перетворюється в класичну формулу для визначення дисперсій середнього арифметичного відхилення.

Закон Пірсона-Джеффріса в наш час широкого використовується в провідних галузях науки. При його застосуванні необхідно визначити ефективні оцінки його параметрів методом максимальної правдоподібності (ММП). Визначимо функцію максимальної правдоподібності L таким чином:

$$L=\prod_{i=1}^{n} y(x_i, a_{VII}, \sigma_{VII}, m) \tag{4}$$

Тоді класичний метод отримання ММП-оцінки параметрів a_{VII}, σ_{VII}, m в формулі (1) зводиться до сумісного розв'язання такої системи рівнянь, запропонованої в [10]:

$$\frac{\partial \ln L}{\partial a_{VII}}=\frac{m}{\mathrm{M}\,\sigma_{VII}^2}\sum_{j=1}^{r} n_j R_j^{-1}(x_j-a_{VII})=0\,; \tag{5}$$

$$\frac{\partial \ln L}{\partial \sigma_{VII}}=\frac{n}{\sigma_{VII}}-\frac{m}{\mathrm{M}\sigma_{VII}^3}\sum_{j=1}^{r} n_j R_j^{-1}(x_j-a_{VII})^2=0\,; \tag{6}$$

$$\frac{\partial \ln L}{\partial m}=n\psi_0-\sum_{j=1}^{r} n_j R_j+\sum_{j=1}^{r} n_j \mathrm{M}_1 R_j^{-1}\left(\frac{x_j-a_{VII}}{\sigma_{VII}}\right)^2=0\,, \tag{7}$$

де, значення $\psi_0=\psi(m+1)-\psi(m+0{,}5)-[2(m-0{,}5)]^{-1}$, $\psi(m)$ – псі-функція, $\mathrm{M}_1=0{,}5m^2(m+1)(m-0{,}5)^{-4}$, n_j – число спостережень в j–му розряді гістограми; $j=1,2,3,\ldots,r$; r – число розрядів; $\sum_{j=1}^{r} n_j=n$; x_j– середина j–того розряду.

Систему диференційних рівнянь (3)–(5) розв'язують методом послідовних наближень, покладаючи у перше наближення таке:

$$a_{VII}{}^1=n^{-1}\sum_{j=1}^{r} n_j x_j\,; \tag{8}$$

$$\sigma_{VII}{}^1=0{,}933\cdot\sqrt{\mu_2}=0{,}933\cdot\sigma\,; \tag{9}$$

$$\sigma = \sqrt{\mu_2}, \quad \mu_2 = \frac{\sum_{i=1}^{n}(x_i - a_{VII})^2 \cdot n_i}{n}; \; m^1 = 4^1. \qquad (10)$$

де μ_2 – другий центральний момент, обчислений для похибок спостережень; σ – їх середнє квадратичне відхилення.

У позначеннях (8–10) оцінки верхній індекс вказує на номер ітерації. У (10) значення $m^1 = 4$ прийнято за рекомендацією сера Г. Джеффріса [10]. В більшості випадків реальні m розподіляються навколо цього значення.

Отримання ММП-оцінки розподілу (1) шляхом розв'язування системи рівнянь (5)–(7) зручно тим, що відразу після його розв'язання можна отримати дисперсії цієї оцінки використовуючи такі формули:

$$-\frac{\partial^2 \ln L}{\partial a_{VII}{}^2} = \sigma_a^{-2}, \quad -\frac{\partial^2 \ln L}{\partial \sigma_{VII}{}^2} = \sigma_\sigma^{-2}, \quad -\frac{\partial^2 \ln L}{\partial m^2} = \sigma_m^{-2} \qquad (11)$$

Дисперсії (11) отримують на основі похідних $\frac{\partial \ln L}{\partial \theta}$ за параметром θ чисельним диференціюванням після того, як система (5)–(7) розв'язана. З цієї причини у рівняннях (5)–(7) не можна робити будь-які скорочення, наприклад, скорочувати коефіцієнти перед знаком суми.

ММП-оцінки, які отримані із рівнянь (5)–(7), є асимптотично ефективними, незалежними і незміщеними, оскільки закон Пірсона-Джеффріса є регулярним і задовольняє всі умови необхідні для отримання таких оцінок.Однак класичний метод оцінки параметрів закону Пірсона-Джеффріса є досить громіздким і займає багато часу. Цей метод розроблено в той час коли ще не було потужних обчислювальних засобів. При наявності таких засобів ММП-оцінку закону (1) можна отримати відразу, навіть без диференціонування функції L в (4). Для цього доцільно скористатись логарифмом функції максимальної правдоподібності для розподілу (1):

$$\ln L = n \cdot \ln\left(\frac{C_{VII} \cdot \Delta}{\sigma_{VII}}\right) - m \cdot \sum_{j=1}^{r} n_j \cdot \ln R_j, \qquad (12)$$

де $n = \sum_{j=1}^{k} n_j$, $k -$ число інтервалів гістограми;

$\Delta -$ ширина інтервалу гістограми;

$$C_{VII} = \Gamma(m+1)\left[\sqrt{2\pi(m-0{,}5)}\cdot\Gamma(m+0{,}5)\right]^{-1};\; R_i = 1+\frac{0{,}5}{\text{M}}\left(\frac{x_i - a_{VII}}{\sigma_{VII}}\right)^2;$$

$$\text{M} = (m-0{,}5)^3 m^{-2}.$$

Прийнявши у (12) в першому наближенні значення параметрів (8)–(10), шукаємо максимум функції (13), змінюючи послідовно оцінки кожного з параметрів спочатку з таким кроком:

$$a_\Delta = \frac{0{,}3\cdot\sigma}{\sqrt{n}};\quad \sigma_\Delta = \frac{0{,}3}{\sqrt{2n}};\quad m_\Delta = \frac{6}{\sqrt{n}} \qquad (13)$$

Обчисливши оцінки відповідні максимуму функції (12) з кроками (13) за кожним із параметрів, аналогічно обчислюємо наступні наближення, зменшивши кроки (13) на третину і так далі, поки не отримаємо остаточні оцінки a_{VII}, σ_{VII}, m параметрів закону Пірсона-Джеффріса (1). Дисперсії цих оцінок в цьому випадку отримують, використовуючи границі нерівностей Рао-Крамера (2-3).

На основі описаного вище методу отримання ефективних оцінок параметрів розподілу (1) створено відповідне програмне забезпечення.

Для вирішення цієї задачі обрано мову програмування С++, оскільки ця мова є однією з найбільш актуальних у наш час. Для створення програмного продукту обрано середовище візуальної розробки програм Borland C++ Builder 6. Це середовище розробки програм є найбільш доцільним для досягнення мети нашого дослідження тому, що це один із найсучасніших інструментів для створення програмних продуктів з графічним інтерфейсом і при цьому створений програмний засіб буде зрозумілим і зручним у використанні навіть для користувачів, які не пов’язані з програмуванням. C++Builder – це сучасний потужний засіб для швидкої розробки додатків (RAD – Rapid Application Development) на C++ під Windows, який підтримує можливість програмування, що ґрунтується на компонентах. Бібліотека візуальних компонентів VCL спрощує розробку програмних засобів завдяки готовим компонентам при використанні яких зменшуються обсяги рутинної роботи. Завдяки візуальному об’єктно-орієнтованому програмуванню була створена технологія, що одержала назву швидка розробка додатків (RAD). Ця технологія характерна для нового покоління систем програмування, до якого відноситься й C++Builder.

За допомогою створеного програмного продукту, можна отримати ефективні оцінки параметрів розподілу. В якості прикладу оцінок параметрів розподілу (1) ми скористались даними лазерних спостережень штучних супутників Землі (ШСЗ) за міжнародною програмою MERIT [11], взяті із роботи [12] і які вказані у табл. 1.

Таблиця 1

Гістограма розподілу похибок «Observation-Calculation» для лазерних спостережень ШСЗ, (розподіл 2, n=575) [3]

№	Середина інтервалів x_i	Частоти n_i	№	Середина інтервалів x_i	Частоти n_i	№	Середина інтервалів x_i	Частоти n_i
1	1,2	1	8	0,5	17	15	-0,2	40
2	1,1	0	9	0,4	22	16	-0,3	29
3	1	0	10	0,3	60	17	-0,4	14
4	0,9	1	11	0,2	81	18	-0,5	8
5	0,8	1	12	0,1	92	19	-0,6	1
6	0,7	1	13	0	109	20	-0,7	1
7	0,6	4	14	-0,1	93			

Визначимо оцінки a_{VII}, σ_{VII}, m параметрів для розподілу (2) за допомогою створеної програми.

Перед початком обчислень необхідно завантажити у програму вектори x_i та n_i, які вказані у табл. 1 і задати вхідні дані у головному вікні: число спостережень n=575, аргумент гамма-функції m=4, ширину інтервалу $\Delta = 0{,}1$, та розмір векторів вхідних даних 20. Після введення цих даних комп'ютер видає результат за 11–12 сек. Результати роботи програми показані на рис. 1.

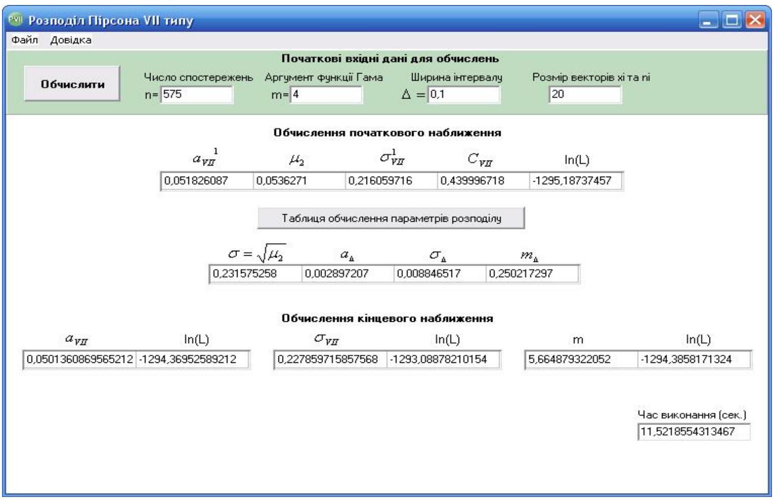

Рис. 1. Форма із результатом роботи програми

Для цього розподілу ми отримали такі ММП-оцінки параметрів:

$$a_{VII} = 0{,}0501360869;$$
$$\sigma_{VII} = 0{,}227859715;$$
$$m = 5{,}664879322.$$

Порівняємо результати обчислень параметрів розподілу (1) з параметрами, обчисленими методом Гауса (табл. 2).

Таблиця 2

Порівняння обчислень параметрів розподілу методами Гауса і Пірсона-Джеффріса

Метод Гауса	**Розподіл Пірсона-Джеффріса**
$a = 0{,}032 \pm 0{,}010$;	$a_{VII} = 0{,}0501 \pm 0{,}009$;
$\sigma = 0{,}230 \pm 0{,}007$;	$\sigma_{VII} = 0{,}228 \pm 0{,}008$;
$m = \infty$.	$m = 5{,}66 \pm 1{,}95$.

Якщо порівняти параметри a і a_{VII} відповідно розподілів Гауса і Пірсона VII типу diag, то впадає в очі різниця 0,0501-0,032=0,0181, яка вдвічі перевищує середню квадратичну похибку визначення параметру a. По параметру σ не спостерігається істотної різниці, проте, найбільш істотна різниця спостерігається для числа m. Для нормального розподілу $m = \infty$. В цьому як раз і полягає основний недолік класичних процедур обробки даних, в яких завжди допускається, що $m = \infty$. Насправді ж реальні розподіли при n>500 досить сильно відхиляються від закону Гауса, оскільки $m = 5{,}66$ дуже далеке від нескінченності.

Підводячи підсумки дослідження можна зробити такі **висновки**:

1. Якщо похибки багаторазових спостережень підкоряються розподілу Пірсона-Джеффріса, то класичні методи обробки спостережень потрібні лише як попередній етап обробки даних. Кінцеві рішення отримують на основі використання вагової функції розподілу похибок «Observation-Calculation».

2. Розроблений в МЕГУ на основі викладеного обчислювального алгоритму програмний продукт у середовищі Borland C++ Builder дозволяє набагато продуктивніше ніж класичним методом отримувати ММП-оцінки закону Пірсона-Джеффріса, що необхідні для реалізації сучасних процедур математичного моделювання.

1. Bessel F. W. Untersuchengen über die Wahrscheinlich keit der Beobachtung fehler / F. W. Bessel // Astronomische Nachrichten. B. 15, 1838, z. 369. **2.** Джунь Й. В. Математическая обработка астрономической и космической информации при негауссовых ошибках наблюдений. Автореферат дис. Доктора физ.-матем. Наук. – К. : ГАО АНУ. 1992. – 46 с. **3.** Dzhun I. V. About make use of Pearson`s Distvibution of Type VII for the Approximation of observation's Errors in Astrometry / I. V. Dzhun // Measurement Techniques: Springer Science + Business Media. Inc. – 1992, vol. 35, № 3, pp. 298–304. Dzhun I. V. Pearson`s Distribution of type VII of the Errors of Satellite Laser Ranging Data / I. V. Dzhun // Kinematics and Physics of Celestial Bodies. – New York : Allerton Press, Inc., 1991, vol.7, pp. 74–84. **5.** Dzhun I. V. Distribution of Errors in multiple large-volume observations / I. V. Dzhun // Measurement Techniques : Springer Sciene + Business Media. Inc. – 2012, vol. 55, № 4, pp. 393–396. **6.** Gazda V. Normal probability Distribution in financial Theory – false Assumption and Consequences / V. Gazda // Department of Economics, University of Economics, Faculty of Businness Economics, Kosice, 1999, p. 5–8. **7.** Хастингс Н. Справочник по статистическим распределениям / Н. Хастингс, Дж. Пикок. – М. : Статистика, 1980. – 95 с. **8.** Dzhun I. V. Comments of Use of the Type VII Pearson Law in Astromentry / I. V. Dzhun, P. V. Novitskii // Kinematics and Physics of Celestial Bodies // Allerton Press. Inc. / New York, 1992, vol. 8. No. 5. p. 78–81. **9.** Джунь И. В. О границах неравенства Рао-Крамера для дисперсий оценок параметров распределения Пирсона VII типа / И. В. Джунь // Кинематика и физика небесных тел. – 1988. – Т. 4. – № 1. – с. 85–87. **10.** Jeffreys H. Theory the Probability / H. Jeffreys. – Oxford. Clarendon Pres, 1998. – 470 p. **11.** Яцків Я. С. Международный проект МЕРИТ. Информ. бюл. Секции астрометрии Астросовета АН СССР / Я. С. Яцків. – К. : Изд-во ГАО АН УССР, 1980. – Вып. 2. – С. 3–9. **12.** Джунь Й. В. Распределение Пирсона VII типа ошибок лазерных наблюдений ИСЗ / Й. В. Джунь // Кинематика и физика небесных тел. – 1991. – Т. 7. – № 3. – С. 82–91.