УДК 528.11

Джунь Й. В., д.ф.-м.н., професор, Карпік С. О., магістр (Міжнародний економіко-гуманітарний університет імені академіка Степана Дем'янчука, м. Рівне)

ЗАКОН РОЗПОДІЛУ ПОХИБОК БАГАТОКРАТНИХ СПОСТЕРЕЖЕНЬ ВЕЛИКИХ ОБСЯГІВ І ОЦІНКА ЇХНІХ ПАРАМЕТРІВ

Анотація. *В статті показано значення і визначні математичні властивості універсального розподілу похибок спостережень великих обсягів – закону Пірсона-Джеффріса, для оцінок параметрів якого запропоновано сучасний програмний продукт у середовищі BorlandC++Builder 6.*
Ключові слова: *програмний продукт, закон Пірсона-Джеффріса, похибки спостереження, великі обсяги.*

Аннотация. *В статье показано значение и определены математические свойства универсального распределения погрешностей наблюдений больших объемов – закона Пирсона-Джеффриса, для оценок параметров которого предложено современный программный продуктов среде Borland C++Builder6.*
Ключевые слова: *программный продукт, закон Пирсона-Джеффриса, погрешности наблюдения, большие объемы.*

Annotation. *The article shows the importance and remarkable mathematical properties of universal distribution of observational errors of large volumes - Pearson–Jeffreys' Law for estimations of parameters of which modern software in the environment BorlandC + + Builder 6 has proposed.*
Keywords: *software product, Pearson–Jeffreys' Law, errors of observation, large volumes.*

Основною особливістю сучасних наукових експериментів є великі або надзвичайно великі обсяги вимірювальної інформації, яка їх забезпечує. Це зумовлено автоматизацією спостережень. Проте математичне моделювання й обробку цих експериментів виконують на основі застарілих уявлень і застарілих методів. Виникає проблема використання сучасних спроможних математичних процедур при аналізі великих обсягів даних. Тому метою із завданням цього дослідження є ознайомлення широкого кола науковців, дослідників, аспірантів, студентів з сучасними поглядами на математичну обробку інформації і з новим фундаментальним законом похибок, які використовують при обробці вибірок великих обсягів, а також із методами оцінки його параметрів.

Відомий англійський математик, геофізик і астроном, кембриджський професор сер Г. Джеффріс ще в 1937 р. зробив висновок, що класичний закон похибок Гауса виявляє свою повну теоретичну і практичну неспроможність за умови, якщо число багатократних спостережень $n>500$. Цю концепцію Джеффріса з великими труднощами було визнано справедливою, оскільки закон Гауса, як фундаментальний принцип математичного моделювання, прекрасно зарекомендував себе протягом більше ніж ста років його практичного використання. Більшість математиків вважали, що саме до закону Гауса мають прямувати похибки спостережень, як до своєї ідеальної граничної форми. Крім того, математизація умов, за яких похибки спостережень прямують до закону Гауса, складає суть центральної граничної теореми теорії ймовірностей. Тому здавалось, що пропозиція Джеффріса знаходиться в дисонансі й із набутим досвідом математичної обробки даних, і з теорією ймовірностей.

Оскільки сер Г. Джеффріс є дуже авторитетним ученим, у НАН України в 70-х рр. минулого століття під керівництвом академіка Е. П. Федорова було здійснено фундаментальну перевірку згаданої вище концепції. Для такої перевірки було залучено 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 істотно негаусів 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).

Крім того, у якій би галузі науки не досліджували похибки спостережень, в усіх випадках їх можна описати магічною формою Джеффріса (1).

Цікавим є походження закону (1). Джеффріс називає цей закон розподілом Пірсона VII типу. Але це не зовсім відповідає істині. Класичний розподіл Пірсона VII типу, який має недіагональну інформаційну матрицю, Джеффріс перетворив 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 до того, як у свій час було визначено закон Гауса.

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

1. Є узагальненням двох найбільш важливих у теорії помилок і аналізі даних розподілів: Гауса, у який формула перетворюється при $m=\infty$, і Стьюдента при значеннях mв (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 таким, що дорівнює ∞. 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]:

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

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

де $\psi'(m)$ – тригама – функція.

Із формули (2) видно, що при $m=\infty$ (закон Гауса) σ_a^2 і σ_σ^2 перетворюється на класичні формули для визначення дисперсій середнього арифметичного і середнього квадратичного відхилень.

Закон Пірсона-Джеффріса набув у наш час широкого використання, здебільшого у провідних галузях науки. При його застосуванні необхідно визначити ефективні оцінки його параметрів на основі методу максимальної правдоподібності (ММП). Для цього 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-оцінок параметрів a_{VII}, σ_{VII}, m в формулі (1) зводять 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 на основі методу послідовних наближень, покладаючи у перше наближення таке:

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

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

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

$$m^1 = 4^1. \tag{10}$$

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

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

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

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

Дисперсії (11) отримують на основі похідних $\frac{\partial \ln L}{\partial \theta}$ за параметром θ чисельного диференціювання 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 отримують 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 [2].

Однак класичний метод отримання оцінок параметрів закону Пірсона-Джеффріса досить громіздкий і займає багато часу. Цей метод розроблено в той час, коли ще не було потужних обчислювальних засобів. За наявності таких засобів ММП-оцінки закону (1) можна отримати відразу, навіть без диференціювання функції L в (4). Для цього доцільно скористатись логарифмом функції максимальної правдоподібності для розподілу (1):

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

де $n=\sum_{j=1}^{k} n_j$, $k-$ число інтервалів гістограми, $\Delta-$ ширина інтервалу гістограми, $C_{VII}=\Gamma(m+1)\left[\sqrt{2\pi(m-0,5)}\cdot\Gamma(m+0,5)\right]^{-1}$,

$R_i=1+\frac{0,5}{\mathrm{M}}\left(\frac{x_i-a_{VII}}{\sigma_{VII}}\right)^2$, $\mathrm{M}=(m-0,5)^3 m^{-2}$.

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

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

Обчисливши оцінки відповідні максимуму функції (12) з кроками (13) за кожним із параметрів, аналогічно обчислюємо такі 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 такому разі 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 RapidApplicationDevelopment (RAD) на C++ під Windows, який підтримує можливість програмування, що ґрунтується на компонентах. Бібліотека візуальних компонентів VCL спрощує розроблення програмних засобів завдяки готовим компонентам, під час використання яких відбувається зменшення обсягів рутинної роботи. Завдяки візуальному об'єктно-орієнтованому програмуванню було створено технологію, що одержала назву «швидка розробка додатків RAD». Ця технологія характерна для нового покоління систем програмування, до якого належить й C++Builder 6.

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], які взято із роботи [12] і вказано у табл. 1. Визначимо оцінки a_{VII}, σ_{VII}, m параметрів для розподілу (2) за допомогою створеної програми.

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**			

Перед початком обчислень необхідно завантажити у програму вектори x_i та n_i, які вказано 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 показано на рис. 1.

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 на основі методу Гауса.

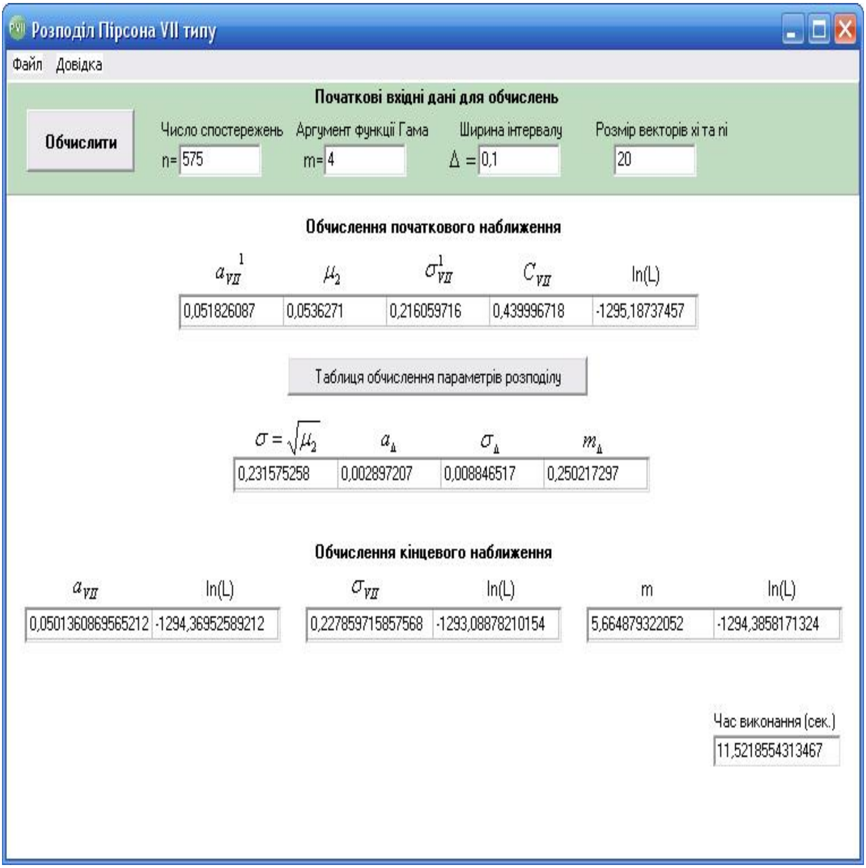

Рис. 1. Форма із результатом роботи програми

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

Унаслідок проведеного 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].

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

1. Bessel F.W. Untersuchengen über die Wahrscheinlichkeit der Beobachtung fahler. // Astronomische Nachrichten. B. 15, 1838, z. 369. **2.** Джунь Й. В. Математическая обработка астрономической и космической информации при негауссовых ошибках наблюдений : автореф. дис. доктора физ.-мат. наук. – К. : ГАО АНУ. 1992. – 46 с. **3.** Dzhun I. V. About make Use of Pearson`s Distribution of Type VII for the Approximation of observation’s Errors in Astrometry // Measurement Techniques: Springer Science + Business Media. Inc. – 1992, vol. 35, № 3, pp.298-304. **4.** Dzhun I. V. Pearson`s Distribution of type VII of the Errors of Satellite Laser Ranging Data.// 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 // Measurement Techniques: Springer Science + Business Media. Inc. – 2012, vol. 55, № 4, pp. 393-396. **6.** Gazda V. Normal probability Distribution in financial Theory – false Assumption and Consequences. // 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., Novitskii P.V. Comments of Use of the Type VII Pearson Law in Astromentry. // Kinematics and Physics of Celestial Bodies.// AllertonPress. Inc./ NewYork, 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 // Кинематика и физика небесных тел. – 1988, т. 4, № 1. – с. 85–87. **10.** Jeffreys H. Theory the Probability. 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.

Рецензент: д.т.н., професор Власюк А. П.