ВІДКРИТТЯ Г. ДЖЕФФРІСА І ЙОГО ЗНАЧЕННЯ ДЛЯ МАТЕМАТИЧНОЇ ОБРОБКИ СУЧАСНИХ ЕКСПЕРИМЕНТІВ BIG DATE

Джунь Й. В.
доктор фізико-математичних наук, професор,
завідувач кафедри математичного моделювання
Міжнародного економіко-гуманітарного університету
імені академіка Степана Дем’янчука
м. Рівне, Україна

Закон похибок в експериментах великого обсягу не відповідає класичним уявленням, оскільки не є гаусовим. Це переконливо показав у своїх роботах знаменитий кембриджський професор Г. Джеффріс, чи сер Гарольд, як його рекомендувала називати леді Джеффріс в листуванні з автором. Тому концепція закону похибок спостережень в експериментах Big Date є основоположною при їх математичній обробці і комп’ютерному програмуванні.

Вперше такі експерименти появились в Англії – це автоматизовані ряди спостережень в Гринвічі на плаваючому зеніт-телескопі Куксона. Два унікальні ряди спостережень на цьому телескопі за період 1927–1931 роках і за 1932–1936 роках мали гігантські на той час обсяги відповідно 4540 і 4810 спостережень [1] і відрізнялись небаченими ексцесами: + 4,14±0,07 і + 6,00±0,06. Як бачимо ці ряди спостережень мали істотно негаусів характер розподілу похибок. Таким чином, саме Англії належить пріоритет у постановці автоматизованих експериментів типу Big Date.

У відповідності з законами великих чисел, із збільшенням числа вимірів, поступово виявляються раніше непомітні особливості досліджуваного процесу, проявляються чіткіше характерні закономірності розподілів похибок. Ці закономірності вперше математично описав сер Гарольд в статтях [1, 2] і у своєму фундаментальному труді «Theory of Probabitity» [3], який витримав у Великобританії 8 перевидань, починаючи з 1939 року. Вивчення цих закономірностей ознаменувалося відкриттям закону похибок спостережень, математично бездоганного у всіх відношеннях. Новий відкритий Джефферісом закон похибок отримав назву закону Пірсона-Джеффріса [4, 5]. На основі цього відкриття Джеффріса в Україні створена «Некласична теорія похибок вимірів» [4, 5], яка відображає найновіші досягнення при обробці експериментів Big Date. Саме такими є сучасні експерименти внаслідок їх автоматизації і комп'ютеризації. Згідно з висновком Джеффріса – при числі спостережень $n > 500$ гіпотеза нормальності їх похибок практично і теоретично є вже неспроможною. Це він переконливо показав в роботі [2], проаналізувавши 6 рядів спостережень штучної зірки з обсягами в межах 500–520. Ці ряди виконані самими К. Пірсоном і його співробітниками [6].

Визначимо суть і значення тих еволюційних процедур, походження яких обумовлено відкриттям Джеффрісом нового, більш універсального закону похибок, які необхідно застосовувати при математичній обробці сучасних експериментів Big Date обсягом $n > 500$. Експерименти такого обсягу, як було показано в результаті аналізу 51 опублікованого ряду за період 1838-1983 років (із загальним обсягом 135838 спостережень), мають, як правило Джеффрісові похибки [4, 5].

У чому ж полягає суть зазначеної вище еволюції методів математичної обробки даних у зв'язку з із зміною парадигми про закон похибок спостережень, що її започаткував Джеффріс? Скажімо, що можливість такої еволюції передбачав ще великий німецький математик К. Ф. Гаус у своїй першій праці по теорії МНК, [7] де він зазначив концепцію нормальності похибок з урахуванням можливості майбутньої еволюції цього методу у вигляді такої формули:

$$f'(x)/\vartheta \cdot f(x) = const,) \tag{1}$$

тут $f'(x)$ – похідна від закону щільності розподілу похибок $f(x)$; похибка $\vartheta = x - a$; a математичне сподівання закону похибок $f(x)$.

Формула (1) виявилася передбачливою і глибокою за сенсом. Саме тому на превеликий жаль, багато із дослідників не зрозуміли її походження і значення, застосовуючи різні евристичні процедури замість того, щоб більш уважно розглянути ліву частину формули (1) у якій якраз і присутня можливість еволюції МНК. Справа в тому, що ліва частина в (1) випливає із застосування Гаусом методу максимальної правдоподібності (ММП), який передбачає винятково нормальності $f(x)$. А вся формула (1) описує лише частковий найбільш бажаний, простий випадок при обробці даних і моделюванні.

Щоб зрозуміти сенс лівої частини формули (1) розглянемо її походження. Вона є результатом мінімізації функції максимальної правдоподібності L в ММП:

$$L = \prod_i^n \; f\,(x_i), \tag{2}$$

де $f(x_i)$ – щільність розподілу імовірності похибки ϑ_i в точці x_i; n – обсяг вибірки.

Логарифмуючи функцію (2) і допускаючи, що L залежить тільки від a, маємо:

$$dln\,L/da = \sum^n [f'(x)/f(x)] = 0. \tag{3}$$

Помноживши чисельник і знаменник у формулі (3) на $\vartheta_i = x_i - a$ отримуємо оцінку $\overline{a}$ параметра $a,$ методом послідовних наближень:

$$\overline{a} = \sum[x_i \cdot \mathrm{P}(\vartheta_i)]/\sum \mathrm{P}(\vartheta_i), \tag{4}$$

де

$$\mathrm{P}(\vartheta_i) = \frac{f'(x_i)}{\vartheta_i} \cdot f(x_i) \tag{5}$$

є ваговою функцією розподілу похибок, (не обов’язково нормальних), яку і використав Гаус у формулі (1). Гаус також попереджував в роботі [8] що «ніхто не може сказати, яким насправді буде закон розподілу похибок спостережень, якщо їх продовжити до нескінченності» [8]. На жаль, ряд високоповажних вчених не оцінили дійсний сенс лівої частини формули (1) і занурилися в евристичний хаос робастних процедур, знехтувати аналітичним методом Гауса для отримання вагової функції при негаусових спостереженнях.

Якщо ми бажаємо використати новий закон похибок з ваговою функцією (5), то необхідно щоб він був в такій же мірі визначений як і розподіл нормальний, тобто, був:

– симетричним, інакше виміри втрачають свій сенс;

– регулярним в межах від – ∞ до + ∞;

– мав незалежності як у закона Гаусса параметри, що означає необхідність діагональності інформаційної матриці Фішера для цього нового розподілу.

Величезна наукова заслуга Джеффріса полягає в тому, що він запропонував новий, більш універсальний ніж розподіл Гауса і математично бездоганний закон похибок (PJ – розподіл), який задовольняє всім трьом вище зазначеним вимогам. Крім звичайних для закону Гауса параметрів – математичного сподівання і дисперсії, джеффрісові похибки мають іще третій, новий, ключовий параметр m, залежний від ексцесу розподілу і тонко реагуючий на так зване «дзижчання» процесу вимірів у часі, чи, інакше, не відчутне нам ритмічне «гудіння» простору вимірів по дисперсії. Якщо до цього добавити, що жоден із відомих трипараметричних розподілів не має діагональної інформаційної матриці, то історична заслуга Джеффріса стає цілком очевидною.

А. Эддінгтон сформулював теорему в якій доказано, що суміш нормальних похибок з однаковим математичним сподіванням і різними дисперсіями неминуче викликає додатній ексцес сумарного розподілу. Більш широко розглянув це питання учень А. Н. Колмогорова, Н. А. Бородачев в книзі [10].

Відносно теорії математичної обробки даних Джеффріс зробив такий основний висновок: «Вирішальним питанням в комбінації спостережень є знання того, чи дійсно розподіли слідують нормальному закону, якщо це не так, то необхідно застосувати інші методи, властиві даному закону» [2]. Цей висновок виявився нині дуже своєчасним, особливо в XXI столітті, коли внаслідок автоматизації і комп'ютеризації вимірів, обсяги вибірок стали значно перевищувати встановлену ним границю в $n > 500$ вимірів.

Джеффрисові похибки мають наступний диференціальний закон щільності [2]:

$$f(x) = \frac{c}{\sigma}\left[1 + \frac{0{,}5}{M}\left(\frac{x-a}{\sigma}\right)^2\right]^{-m}, \tag{6}$$

де стала $c = \left[(2m-1)^{0{,}5} \cdot B\left(m + \frac{1}{2}, \frac{1}{2}\right)\right]^{-1}$; $B(w, z)$ – бета – функція; $M = (m - 0{,}5)^3 \cdot m^{-2}$; $a, \sigma -$ відповідно математичне сподівання і міра розсіювання; m – ключовий параметр розподілу (6).

Використовуючи формулу (5, 6) отримуємо вагову функцію джеффрісових похибок ϑ_i:

$$P(\vartheta i) = \left[\left(\frac{m-0{,}5}{m}\right)^3 \cdot \sigma^2 + \frac{\vartheta i^2}{2m}\right]^{-1}, \tag{7}$$

де $\sigma, m -$ММП – оцінки параметрів розподілу (6).

Оскільки вага спостереження – це його обернена дисперсія, то $P(\vartheta i)$ в (7) – це вага спостереження x_i, джеффрісова похибка якого ϑ_i.

Перерахуємо, що може дати використання розподілу (6) і його вагової функції (7) для вдосконалення обробки спостережень Big Date:

1. Вагова функція (7) дозволяє реалізувати некласичну теорію МНК за умови:

$$\frac{f'x_i}{\vartheta \cdot f(x_i)} = P(\vartheta_i) \neq const, \tag{8}$$

при вимозі $\sum P(\vartheta_i) \cdot \vartheta_i^2 = min$.

2. Отримувати ефективні оцінки зваженої середньої, коли обсяги вибірок $n > 500$.

3. Нормалізувати джеффрісові похибки у випадках застосування критеріальних процедур, які грунтуються на законі Гауса, шляхом використання такого оператора:

$$x_{\text{н}i} = x_i\sqrt{P(\vartheta_i),} \tag{9}$$

де $x_{\text{н}i}$ – нормалізовані спостереження; $P(\vartheta_i)$ – вага спостереження x_i яку обчислюють по формулі (7).

Таким чином, формула (9) дає можливість успішно використовувати класичні програмні пакети критеріальних процедур у випадку джеффрісових похибок.

4. Приводити до стаціонарного вигляду нестандартні по дисперсії динамічні ряди при їх спектральному аналізі.

5. Діагностувати якість експеримента Big Date на предмет попадання оцінки m розподілу(6) в границі $3 \leq m \leq 5$, що, по Джеффрісу є свідоцтвом чисто випадкового характеру похибок ϑ_i [2].

Описані вище нові можливості некласичної теорії похибок вимірювань створені в Міжнародному економіко-гуманітарному університеті на основі використання відкритого Г. Джеффрісом закону (6), відображають надзвичайно потрібні зміни в еволюції методів обробки автоматизованих спостережень Big Date, які є типовими в XXI столітті.

Література:

1. Jeffreys H. The Law of Errors in the Greenwich Variation of Latitude observations. *Mon. Not. Of the RAS*. 1939. Vol. 99. № 9. Pp. 703–709.
2. Jeffreys H. The Law of Errors and the Combinations of Observations. *Philos. Trans. Roy. Soc. London*, 1937, ser. A. № 237. Pp. 231–271.
3. Jeffreys H. Theory of Probability. Sec. Edition. *Oxford University,* 1998. 470 p.
4. Джунь И. В. Неклассическая теория погрешностей измерений. Ровно : *Естеро*. 168 с.

5. Dzhun I. V. Non – Classical Theory Measurements Errors. USA : *Amazon.* 200 p.
6. Pearson, R. On the Mathematical Theory of Errors of Judgment with special Reference to the Personal Equation. *Philosophical Transactions of the Royal Society of London.* 1902. Ser, A., 198, 235–236.
7. Gauss C. F. Theoria motus corporum celestium in sectionibus conicis Solem ambientium. *Hamburgi*. 1809.
8. Гаусс К. Ф. Избранные геодезические сочинения. Том 1. Способ наименьших квадратов / под. ред. Г. В. Багратуни М. : *Геодезиздат*, 1975. 262 с.
9. Eddington A. S. Notes on the Method of Least-Squares. The Proceedings of the Phisical Society, 1933, vol. 45. Part 2, № 247. Pp. 135–365.
10. Бородачев Н. А. Основные вопросы теории точности производства. Бородачев Н. А. / под ред. А. Н. Колмогорова. М. – Л. : Изд. АН СССР, 1950. 360 с.