

# Методи Визначення Класу Розподілу Екстремальних Значень Результатів Фізичного Експерименту

Кирило Сазонов  
кафедра прикладної математики  
Харківський національний університет  
радіоелектроніки  
Харків, Україна  
kir1218@hotmail.com

Володимир Кобзев  
кафедра прикладної математики  
Харківський національний університет  
радіоелектроніки  
Харків, Україна  
volodymyr.kobziev@nure.ua

## Methods of Determining Extreme Values Class of Physical Experiment Distribution Results

Kyrylo Sazonov  
Applied Mathematics Department  
Kharkiv National University of Radioelectronics  
Kharkiv, Ukraine  
kir1218@hotmail.com

Volodymyr Kobziev  
Applied Mathematics Department  
Kharkiv National University of Radioelectronics  
Kharkiv, Ukraine  
volodymyr.kobziev@nure.ua

**Анотація**—Наведені постулат стійкості розподілів випадкових величин та заснований на ньому принцип класифікації граничних розподілів. Викладені два методи визначення класу граничного розподілу на основі хвостового індексу, які проілюстровані результатами обробки значень лептону pT.

**Abstract**—The postulate of stability of distributions of random variables and the principle of classification of boundary distributions based on it is given. Two methods for determining the class of the boundary distribution on the basis of the tail index are presented, which are illustrated by the results of processing the values of the lepton pT.

**Ключові слова**—фізичний експеримент; екстремальні значення; класи граничних розподілів; хвостові індекси

**Keywords**— physical experiment; extreme values; classes of boundary distributions; tail indices

### I. ВСТУП

Проведення фізичних експериментів, спрямованих на дослідження характеру протікання різних процесів чи уточнення значень деяких величин, найчастіше базується на підтвердженні певних теоретичних гіпотез на основі опрацювання числових даних. Обмежений обсяг таких даних зумовлений практичною неможливістю або надмірною вартістю проведення великої кількості експериментів. Ці чинники пояснюють неможливість використання асимптотичних методів опрацювання даних, отриманих експериментальним шляхом.

Найчастіше аналізовані числові дані розглядаються як випадкові величини з деякою функцією розподілу. Разом з середніми значеннями досліджуваних величин чималий практичний інтерес мають найменші та найбільші (екстремальні) значення у вибірках фіксованого об'єму. Саме вони містять важливі відомості про граничні прояви досліджуваних явищ чи процесів.

### II. ПОСТУЛАТ СТІЙКОСТІ ТА КЛАСИ РОЗПОДІЛІВ ЕКСТРЕМАЛЬНИХ ЗНАЧЕНЬ

Розглянемо серію з  $N$  вибірок однакового об'єму  $n$ , узятих з генеральної сукупності, що має розподіл  $F(x)$  випадкової величини  $X$ . В кожній вибірці є найбільше значення. Розподіл в  $Nn$  спостереженнях буде прагнути до тієї ж границі, що й розподіл найбільшого значення в вибірці об'єму  $n$ , за умови, що ця границя існує. Виходячи з цього, граничний розподіл повинен бути таким, щоб взяте з будь-якої вихідної вибірки найбільше значення мало такий же граничний розподіл - це і є постулат стійкості розподілу випадкових величин.

Лінійне перетворення не впливає на вигляд розподілу величин, тому ймовірність того, що найбільше значення менш ніж  $x$ , повинна дорівнювати ймовірності лінійної функції від  $x$  [1]:

$$F^n(x) = F(a_n x + b_n) \quad (1)$$

де параметри  $a_n$  і  $b_n$  є деякими функціями від  $n$ .



Рівняння (1) вперше використовував Фреше в якості математичного формулювання постулату стійкості розподілу величин.

У роботі [1] встановлено, що існує три типи граничних розподілів, для кожного з яких притаманна специфічна поведінка найбільших абсолютних значень випадкової змінної. Перший з встановлених класів охоплює групу розподілів вихідних величин експоненціального типу. Другий клас визначений для розподілів типу Коші з важкими хвостами, причому вихідні змінні необмежені для найбільших значень знизу, для найменших значень згори, чи в обох напрямках. Третій клас справедливий для розподілів величин, які обмежені згори чи знизу.

Крім того, відомо [1], що лінійна функція випадкової величини у кожному з трьох екстремальних розподілах також є екстремальним значенням.

Узагальнений розподіл екстремальних значень являє собою узагальнення розподілів Гумбеля, Фреше та Вейбулла, його застосовують для наближеного модулювання максимумів кінцевих послідовностей випадкових величин.

Вважають, що випадкова величина  $X$  розподілена за узагальненим законом, якщо він представлений функцією:

$$F(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\}, \quad (2)$$

де:  $\mu$  – параметр розміщення,  $\sigma$  – параметр масштабу,  $\xi > 0$  – параметр форми.

За допомогою вказаних параметрів, а також константи Ейлера  $\gamma$  та гама-функції  $g_k = \Gamma(1 - k\xi)$ , можна визначити математичне очікування, дисперсію, медіану, моду та коефіцієнт ексцесу узагальненого розподілу екстремальних значень.

Функція щільності задається виразом:

$$f(\xi, \mu, \sigma) = e^{-\left(1 + \xi \left(\frac{x - \mu}{\sigma}\right)^{-1/\xi}\right)} \cdot \frac{1}{\sigma} \left(1 + \xi \left(\frac{x - \mu}{\sigma}\right)\right)^{-(1/\xi)-1}.$$

Логарифмічна функція розподілу має два вигляди.

При  $\xi \neq 0$ :

$$\ln l(\xi, \mu, \sigma) = -n \ln(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \ln \left[ 1 + \xi \left(\frac{x_i - \mu}{\sigma}\right) \right] - \sum_{i=1}^n \ln \left[ 1 + \xi \left(\frac{x_i - \mu}{\sigma}\right) \right]^{-1/\xi}.$$

Для  $\xi = 0$ :

$$\ln l(\mu, \sigma) = -n \ln(\sigma) - \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma}\right) - \sum_{i=1}^n \exp \left\{ - \left(\frac{x_i - \mu}{\sigma}\right) \right\}.$$

Найкраща оцінка отримується за допомогою методу максимальної правдоподібності в результаті максимізації обох функцій правдоподібності, після чого вибирається максимальний результат.

### III. МЕТОДИ ВИЗНАЧЕННЯ КЛАСУ РОЗПОДІЛУ

Практичне визначення класу граничного розподілу використовує оцінку «важкості» його хвостів за кінцевою вибіркою спостережень випадкової величини.

Існує багато методів оцінки хвостових індексів: оцінка відношення [2], момент-оцінка [3], УН-оцінка [4], DPR оцінка [5] й інші. Розглянемо оцінки Хілла та Пікандса, як найбільш прості у використанні.

Нехай  $X_1, X_2, \dots, X_n$  – упорядкована за спаданням вибірка незалежних спостережень за  $\alpha$ -стійкою випадковою величиною  $X, n \in N$  :

$$X_1 \geq X_2 \geq \dots \geq X_n.$$

Параметр  $\gamma = 1/\alpha$ , де  $\alpha$  – хвостовий індекс, називається індексом екстремального значення й визначає форму хвоста розподілу випадкової величини.

Оцінка Хілла [6] є спроможною оцінкою, основана на  $k+1$  верхній статистиці, величина якої знаходиться як:

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^k (\ln X_{(i)} - \ln X_{(k+1)}), \quad (3)$$

де  $k \in [2; n/2]$ .

Так як  $X_1, X_2, \dots, X_n, n \in N$ , незалежні спостереження, то оцінка Хілла  $H_{k,n}$  спроможна для параметру  $\gamma = \alpha^{-1}$  у наступному сенсі: існує послідовність

$$k, k \rightarrow \infty, k/n \rightarrow 0, \text{ при } n \rightarrow \infty$$

така, що має місце:

$$H_{k,n} \xrightarrow{p} \gamma.$$

На практиці точність оцінки сильно залежить від вибору значення  $k$ . Спосіб знаходження оцінки полягає в побудові графіку:

$$\{H_{k,n}(k) : k = 2, \dots, n/2\}.$$

Оцінку  $k$  обирають з інтервалу значень  $[k_-, k_+]$ , на якому функція  $\{H_{k,n}(k)\}$  демонструє сталість.

В якості оцінки для індексу екстремальних значень  $\gamma$  розглядається статистика:

$$\bar{H}_n = \frac{1}{k_+ - k_- + 1} \sum_{i=k_-}^{k_+} H_{i,n},$$



яка незміщеним чином оцінює  $\gamma$ , тобто  $E(\bar{H}_n) = \gamma$ .

Значення хвостового індексу дорівнює:

$$\alpha = \bar{H}_n^{-1}.$$

Індекс екстремального значення за методом Пікандса, згідно з [6], визначається як:

$$\hat{\gamma}p(k) := (\ln 2)^{-1} \frac{X_{(n-k)} - X_{(n-2k)}}{X_{(n-2k)} - X_{(n-4k)}}. \quad (4)$$

У цьому методі використовується упорядкування спостережень за зростанням  $X_1 \leq X_2 \leq \dots \leq X_n$ , потім як і в методі Хілла будується графік:

$$\{\hat{\gamma}p(k): k = 2, \dots, n/4\}.$$

Після цього знаходиться стала ділянка  $[k_-, k_+]$  і за нею визначається індекс екстремального значення  $\gamma$  як середнє значення елементів цього інтервалу, а хвостовий індекс:

$$\alpha = \gamma^{-1}.$$

Коли  $\alpha = 2$  розподіл є нормальним, при  $0 < \alpha < 2$  розподіл має важкі хвости та витягнутість. У випадку, коли  $\alpha > 3$ , розподіли не мають властивості замкнутості відносно композиції, однак, вони мають важливу властивість збереження характеру убавання щільності розподілу величин при композиції.

#### IV. ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ

Викладені методи застосовані до набору даних значень лептону рТ з USI репозиторію (фундаментальні частки, які не приймають участь у сильній взаємодії) обсягом  $N = 500$ ,  $n = 30$ .

За допомогою пакету STATISTICA встановлено, що за критеріями Колмогорова-Смірнова та Андерсона-Дарлінга розподіл наявних даних найкраще відповідає на рівні значущості 0,95 узагальненому розподілу екстремальних значень. Розраховані параметри та основні статистичні характеристики цього розподілу, емпірична функція наведена на Рис.1.

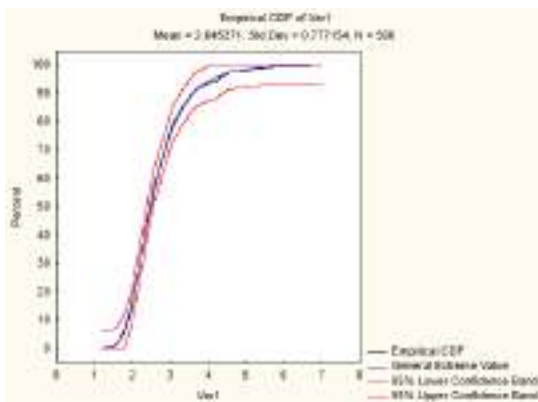


Рис.1 – Графік емпіричної функції розподілу для значення лептону рТ

На Рис. 2 наведені графіки, стабільні ділянки яких дозволяють скористатися формулами (3) та (4) для розрахунків хвостових індексів.

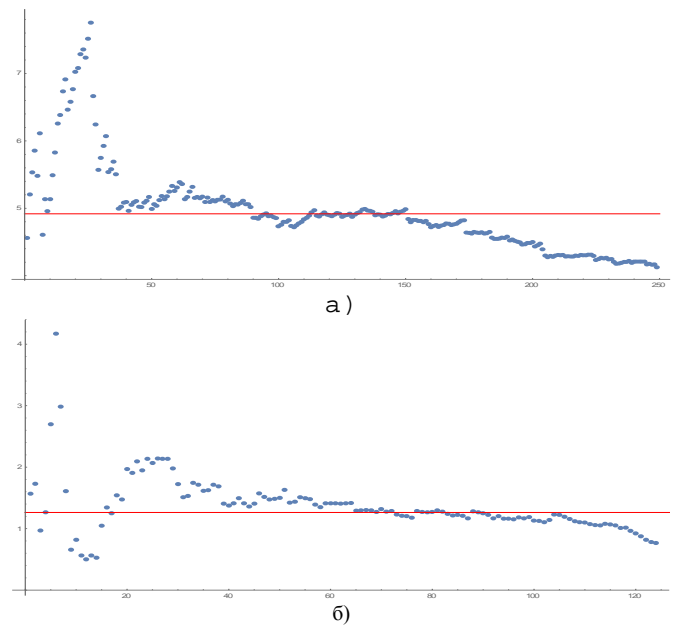


Рис.2 – Оцінювання хвостового індексу для значення лептону рТ.  
а) – метод Хілла, б) - метод Пікандса

При використанні методу Хілла значення сталих інтервалів для лептону рТ  $[k_- = 110, k_+ = 150]$ . Значення хвостового індексу  $\alpha = 4,9182$ . За методом Пікандса індекс екстремального значення  $\gamma = 1,2632$ , значення хвостового індексу  $\alpha = \gamma^{-1} = 0,7916$ . Це підтверджує гіпотезу про приналежність наших даних до третього граничного розподілу екстремальних значень.

#### V. ВИСНОВКИ

Розглянуті методи дозволяють визначати класи розподілу екстремальних значень у наборах експериментальних даних. Їх трудомісткість та точність залежать як від об'єму оброблюваних даних, так і від довжини ділянки, на якій відповідні статистики демонструють сталу поведінку.

#### ЛІТЕРАТУРА REFERENCES

- [1] Гумбель Э. Статистика экстремальных значений. – М.: Мир, 1965. – 450 с.
- [2] Goldie C.M., Smith R.L. Slow variation with remainder: theory and applications.// Quart.J.Math. Oxford. 38. 1987. P. 45–71.
- [3] Dekkers A.L.M., Einmahl J.H.J., de Haan L. A moment estimator for the index of an extreme-value distribution.// Annals of Statistics 17 1989. P. 1833–1855.
- [4] Berline A., Vajda I., van der Maulen E.C. About the asymptotic accuracy of Barron density estimates.// IEEE Trans. Inf. Theory. № 44. 1998. P. 999–1009.
- [5] Davydov Yu., Paulauskas V., Rackauskas A. More on P-stable convex sets in Banach spaces// J.Theoret. Probab. 13. no.1. P. 39–64. 2000.
- [6] Родионов И. В. Статистический анализ и проверка гипотез о распределении экстремумов временного ряда [Текст]: дис. канд. физико-математических наук: защищена 15.01.13: утв. 27.09.13 / В. И. Питербарг. — М., 2013. — 70 с.

