Український математичний вісник
Том 17 (**2020**), № 2, 196 – 214
https://doi.org/10.37069/1810-3200-2020-17-2-4

Арктангенс-регресія та оцінювання параметрів розподілу Коші

Іван Г. Крикун

(Представлена І. І. Скрипніком)

Анотація. В роботі побудовані оцінки для знаходження параметрів нелінійної регресії між величинами X та Y у випадку регресійної функції арктангенс. Отримані оцінки використано для оцінювання невідомих параметрів розподілу Коші. Проведено комп'ютерне моделювання та порівняно отримані оцінки з наявними квантильними оцінками, оцінками максимальної вірогідності та деякими іншими. Отримано довірчі інтервали параметрів розподілу Коші.

2010 MSC. 62E17, 62J12.

Ключові слова та фрази. Нелінійна регресія, розподіл Коші, оцінювання параметрів розподілу, довірчий інтервал.

Вступ

На практиці та в теорії часто зустрічаються залежності між деякими величинами X та Y, які описуються рівняннями нелінійної регресії – поліноміальна, показникова, гіперболічна, логарифмічна. Водночас автору не зустрівся розгляд арктангенсіальної регресії, хоча властивості функції арктенгенс є зручними для застосування на практиці: неперервність, диференційованість, монотонність, обмеженість, визначеність на всій осі. Тож природнім є припущення про доцільність розгляду арктангенсіальної регресії. Приклади її можливого застосування наведено в пропозиції 4.

Стандартна техніка знаходження параметрів арктангенс-регресії методом найменших квадратів не дає можливості знайти ці оцінки в явному вигляді, але цю модель, після звичайної для нелінійного регресійного аналізу нелінійної заміни, вдається звести до лінійної, звідки і отримати оцінки параметрів арктангенс-регресії.

Стаття надійшла в редакцію 30.09.2019

В подальшому тексті для позначення функцій "тангенс" і "арктангенс" використовуються стандартні LaTeX-івські функції $\tan x$ та $\arctan x$ замість прийнятих у вітчизняній літературі позначень $\operatorname{tg} x$ і $\operatorname{arctg} x$.

Нагадаємо, що випадкова величина має розподіл Коші з параметрами (a, γ) (де $\gamma > 0$) якщо її щільність має вигляд [1, стор. 123]

$$f(x) = \frac{1}{\pi} \frac{\gamma}{(x-a)^2 + \gamma^2}, \quad -\infty < x < \infty,$$

а функція розподілу, відповідно,

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{x-a}{\gamma}.$$

Параметр a називають параметром розташування, а параметр γ – параметром масштабу.

Відомо [2, с. 112], що розподіл Коші не має ні математичного сподівання, ні моментів вищих порядків. Даний розподіл належить до розподілів з „важкими хвостами“, для яких не виконується закон великих чисел. Зокрема, середнє арифметичне випадкових величин, розподілених за законом розподілу Коші, також є випадковою величиною, розподіленою за законом розподілу Коші з тими ж параметрами розташування і масштабу [2, с. 114].

В літературі розглянуто кілька підходів до знаходження параметрів розподілу Коші, проте, в силу вищезазначених властивостей розподілу Коші, в загальній ситуації ці підходи дають досить неточні результати та, як правило, пов'язані із досить складними обчисленнями.

Найпростішим підходом до знаходження параметру розташування a є метод вибіркових квантилів, згідно з яким параметр розташування a оцінюється через медіану впорядкованої вибірки [1, с. 123]

$$a^* = x_{0,5} \tag{0.1}$$

або через півсуму вибіркових квантилів порядку p і $1-p$, наприклад 0,75 і 0,25, тобто [2, с. 115]

$$a^* = \frac{x_{0,75} + x_{0,25}}{2}. \tag{0.2}$$

Для параметру масштабу γ однією з простих є оцінка через вибіркові квантилі порядку p і $1-p$, наприклад 0,75 і 0,25, [2, с. 115]

$$\gamma^* = \frac{x_{0,75} - x_{0,25}}{2}. \tag{0.3}$$

Оцінки параметрів розподілу Коші методом максимальної вірогідності знайти в явному вигляді не вдається через складності обчислень. Так, в роботі [3] показано, що у невиродженій ситуації (тобто коли менш, ніж половина вибіркових значень однакові) оцінка максимальної вірогідності для обох невідомих параметрів розподілу Коші існує і єдина та знаходиться як розв'язок деякої системи нелінійних рівнянь. Цей підхід також розглянуто та реалізовано для випадків $n = 3$ і $n = 4$ (що виявилось нелегкою задачею в докомп'ютерний час) в [4]. Було розроблено різні способи знаходження розв'язку цієї системи, як правило з використанням послідовних наближень. Зокрема, в [5, Пример 18.9, с. 76] для випадку $\gamma = 1$ запропонована ітераційна процедура для послідовного наближення оцінки максимальної вірогідності параметру розташування a, а в [6] подібна ітераційна процедура запропонована для загального випадку (для обох невідомих параметрів розподілу).

Оцінки параметрів розподілу Коші методом моментів не знаходять, оскільки цей розподіл не має скінченних моментів. Проте в [7] для випадку $a = 0$ запропоновано для знаходження оцінки параметру масштабу γ використовувати узагальнений метод моментів і знайдено оцінки через моменти порядку k, де $0 < k < 1$. Деякі інші способи оцінювання параметрів розподілу Коші, отримані з їх використанням результати та більш детальний огляд проблематики можна знайти в роботах [8–13].

На практиці розподіл Коші (зокрема його параметри) застосовується в економіці для моделювання змін курсу цінних паперів [7] та для оцінки наслідків стихійних лих [14], у фізиці в задачах пошуку джерела випромінювання [2, п. 9, с. 114] та в задачах, пов'язаних із обробкою сигналів [15].

Робота організована таким чином: в наступному розділі знайдено оцінки параметрів арктангенсіальної регресії. В розділі 2 розглянуто питання про пошук за допомогою арктангенс-регресії оцінок для невідомих параметрів розподілу Коші. В розділі 3 проведено комп'ютерне моделювання розподілу Коші з заданими параметрами розташування і масштабу та порівняно оцінки параметрів, отримані за формулами розділу 2, з оцінками, отриманими іншими способами. Висновки щодо застосування отриманих оцінок та приклади можливого застосування такої регресії в теорії та в практичних задачах розглянуто в розділі 4. В розділі 5 наведено довірчі інтервали для параметрів розподілу Коші. В додатку наведено графіки арктангенс-регресії для першої із розглянутих в розділі 3 вибірок.

1. Арктангенсіальна регресія

В подальшому розглядатимемо n спостережень двовимірної вибірки виду $(x_i; y_i)$.

В загальному вигляді арктангенсіальна залежність між деякими величинами X та Y має вигляд

$$Y = C \arctan(AX + B) + D, \tag{1}$$

де A, B, C, D – параметри регресії.

Нелінійна регресійна модель з чотирма параметрами є складною. Але її можна спростити, вважаючи (виходячи з теоретичних або практичних міркувань) масштабні параметри регресії C і D відомими. Це може бути один з таких випадків:

$$Y = \frac{2}{\pi} \arctan(AX + B); \tag{2}$$

$$Y = \frac{1}{2} + \frac{1}{\pi} \arctan(AX + B). \tag{3}$$

Для знаходження регресійної залежності між Y та X будемо використовувати модель (3).

Оцінки невідомих параметрів регресії будемо знаходити методом найменших квадратів. Застосування МНК напряму призводить до знаходження розв'язку занадто складних систем, тож зведемо модель (3) до лінійної регресії за допомогою заміни

$$z = \tan\Big(\pi(y - \frac{1}{2})\Big). \tag{4}$$

Тоді арктангенс-регресія виду (3) зводиться до лінійної моделі

$$Z = AX + B, \tag{5}$$

оцінки коефіцієнтів якої знаходяться за добре відомими формулами оцінок коефіцієнтів лінійної регресії, які і дають формули нижче.

Пропозиція 1. *Оцінки невідомих параметрів регресії арктангенс-регресії виду (2) мають вигляд*

$$A^* = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} x_i z_i - \frac{1}{n}\sum\limits_{i=1}^{n} x_i \frac{1}{n}\sum\limits_{i=1}^{n} z_i}{\frac{1}{n}\sum\limits_{i=1}^{n} x_i^2 - \Big(\frac{1}{n}\sum\limits_{i=1}^{n} x_i\Big)^2}, \tag{6}$$

$$B^* = \frac{1}{n}\sum_{i=1}^{n} z_i - A^* \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{7}$$

Зауваження 1. Аналогічним чином, використовуючи заміну

$$u = \tan\left(\frac{\pi}{2}y\right),$$

арктангенс-регресія виду (2) зводиться до лінійної моделі

$$U = AX + B,$$

коефіцієнти якої знаходяться за формулами (6)-(7) із заміною величин z на відповідні u.

2. Оцінка параметрів розподілу Коші

Постановка задачі. Розглянемо вибірку з n спостережень випадкової величини X, що розподілена за законом розподілу Коші з невідомими параметрами (a, γ). Потрібно знайти оцінки цих параметрів.

Пропозиція 2. *Будемо шукати оцінку невідомих параметрів розподілу Коші використовуючи арктангенс-регресію і формули для оцінок її параметрів з попереднього розділу.*

Для цього виконаємо наступну процедуру:

1. *Впорядкуємо вибірку за зростанням (тобто побудуємо варіаційний ряд)*
$$x_1, x_2, ..., x_n. \tag{8}$$

2. *Кожному x_i з (8) поставимо у відповідність числа y_i ($i = 1, ..., n$) (зміщену накопичену частоту) за формулою*
$$y_i = \frac{i}{n} - \frac{1}{2n}. \tag{9}$$

3. *Для отриманої двовимірної вибірки $\{(x_i; y_i)\}$ $i = 1, ..., n$ побудуємо арктангенс-регресію виду (3). Заміною (4) зведемо її до лінійної моделі (5), оцінки коефіцієнтів якої A^* і B^* знайдемо за формулами (6)–(7).*

4. *Шляхом очевидних перетворень можемо отримати шукані оцінки невідомих параметрів розподілу Коші*
$$a^* = -\frac{B^*}{A^*}; \tag{10}$$
$$\gamma^* = \frac{1}{A^*}. \tag{11}$$

Зауваження 2. В формулі (9) використана накопичена частота була зменшена на величину $\frac{1}{2n}$ для уникнення занадто великих значень тангенсу, обчисленого за формулою (4).

Зауваження 3. Виходячи із властивостей розподілу Коші (його "важкі хвости", тобто той факт, що у випадкової величини, розподіленої за розподілом Коші з досить великою ймовірністю з'являються великі за модулем значення) можна припустити, що для отримання кращих оцінок варто відкидати найбільші та найменші значення з варіаційного ряду (8). Моделювання в наступному розділі підтверджує це припущення.

3. Моделювання розподілу Коші, оцінка його параметрів за допомогою арктангенсіальної регресії та порівняння з оцінками, отриманими іншими способами

Пропозиція 3. *В програмі MathCad 15 (функція* $\mathbf{rcauchy}(n, a, \gamma)$*) змодельовано* 500 *спостережень випадкової величини, розподіленої за законом розподілу Коші з параметрами* $a = 2$ *та* $\gamma = 5$*. Для показовості було зроблено 4 таких вибірки, кожна з них оброблена за процедурою, що була запропонована в пропозиції 2.*

Для порівняння оцінок, отриманих різними способами також було змодельовано вибірки з $n = 50$ *та* $n = 5000$ *спостережень випадкової величини, розподіленої за законом розподілу Коші з тими ж параметрами.*

Далі за формулами (10)–(11) було знайдено оцінки параметрів розподілу, при цьому для кожної з отриманих вибірок було використано один із наведених нижче способів обробки, які в подальшому будемо коротко називати:

Спосіб №1 *– використовуються всі змодельовані значення;*
Спосіб №2 *– відкинуто 1% найбільших і 1% найменших значень;*
Спосіб №3 *– відкинуто по 10% найбільших і найменших значень;*
Спосіб №4 *– відкинуто по 25% найбільших і найменших значень.*

Для кожної вибірки і для кожного з наведених способів відбору даних за знайденими за формулами (6)–(7) оцінками параметрів арктангенс-регресії для оцінки адекватності отриманої моделі обчислимо коефіцієнт детермінації R^2 (за всіма спостереженнями, тобто з врахуванням відкинутих вибіркових значень) за звичайною формулою

$$R^2 = 1 - \frac{\sum\limits_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{\sum\limits_{i=1}^{n}\left(y_i - \overline{y}\right)^2},$$

де $\overline{y}$ – це середнє арифметичне за всією вибіркою, а $\widehat{y}_i$ – це i-те розраховане за рівнянням моделі значення величини y.

Для порівняння точності оцінок в залежності від обсягу вибірки аналогічно було опрацьовано ще дві вибірки: з 50 та з 5000 реалізацій випадкової величини, розподіленої за законом розподілу Коші з тими ж параметрами $a = 2$ і $\gamma = 5$.

Результати чисельного моделювання наведено у таблицях нижче. В цих таблицях:
— стовпець "Спосіб №" — це спосіб відбору даних згідно пропозиції 3;
— стовпець "R^2" — це коефіцієнт детермінації арктангенс-регресії з отриманими у відповідних рядках параметрами;
— рядок "Квантильні оцінки" — це квантильні оцінки параметрів за формулами (0.1)–(0.3) (тут знайдено дві оцінки параметру a – за формулами (0.1) та (0.2), та одна оцінка параметру γ – за формулою (0.3); для компактності всі ці оцінки наведені в одній таблиці, тому у стовпчику a^* з'являється два значення, так само як і в стовпчику R^2 – коефіцієнт детермінації розраховується для двох різних оцінок параметру a і однієї й тієї ж оцінки параметру γ);
— рядок "Оцінки ММВ" — це оцінки параметрів за методом максимальної вірогідності, отримані шляхом чисельного розв'язання (в Mathcad 15) відповідної системи рівнянь [4, система (2)].

В таблицях напівжирним шрифтом виділено оцінки параметрів розподілу Коші, що найкраще відповідають істинним значенням $a = 2$ і $\gamma = 5$.

Таблиця 1. Результати моделювання параметрів $a = 2$ і $\gamma = 5$ розподілу Коші за вибіркою в 500 спостережень № 1

Спосіб №	a^*		γ^*	R^2	
1	-0,270		5,372	0,9544	
2	2,634		5,442	0,9711	
3	1,353		5,331	0,9982	
4	1,311		**5,036**	**0,9982**	
Квантильні оцінки	**1,571**	0,985	5,348	0,9975	0,9957
Оцінки ММВ	1,389		5,13	0,882	

Таблиця 2. Результати моделювання параметрів $a = 2$ і $\gamma = 5$ розподілу Коші за вибіркою в 500 спостережень № 2

<table>
<tr><th>Спосіб №</th><th colspan="2">a^*</th><th>γ^*</th><th colspan="2">R^2</th></tr>
<tr><td>1</td><td colspan="2">7,930</td><td>8,05</td><td colspan="2">0,679</td></tr>
<tr><td>2</td><td colspan="2">5,678</td><td>10,901</td><td colspan="2">0,848</td></tr>
<tr><td>3</td><td colspan="2">2,197</td><td>5,405</td><td colspan="2">**0,9997**</td></tr>
<tr><td>4</td><td colspan="2">2,232</td><td>**5,256**</td><td colspan="2">0,9996</td></tr>
<tr><td>Квантильні оцінки</td><td>2,24</td><td>**2,059**</td><td>5,429</td><td>0,999</td><td>0,999</td></tr>
<tr><td>Оцінки ММВ</td><td colspan="2">2,199</td><td>5,341</td><td colspan="2">0,977</td></tr>
</table>

Таблиця 3. Результати моделювання параметрів $a = 2$ і $\gamma = 5$ розподілу Коші за вибіркою в 500 спостережень № 3

<table>
<tr><th>Спосіб №</th><th colspan="2">a^*</th><th>γ^*</th><th colspan="2">R^2</th></tr>
<tr><td>1</td><td colspan="2">4,632</td><td>3,228</td><td colspan="2">0,876</td></tr>
<tr><td>2</td><td colspan="2">4,845</td><td>6,573</td><td colspan="2">0,936</td></tr>
<tr><td>3</td><td colspan="2">3,201</td><td>5,556</td><td colspan="2">0,9944</td></tr>
<tr><td>4</td><td colspan="2">2,806</td><td>4,837</td><td colspan="2">**0,9976**</td></tr>
<tr><td>Квантильні оцінки</td><td>**2,555**</td><td>3,20</td><td>5,054</td><td>0,996</td><td>0,995</td></tr>
<tr><td>Оцінки ММВ</td><td colspan="2">2,605</td><td>**4,966**</td><td colspan="2">0,881</td></tr>
</table>

Таблиця 4. Результати моделювання параметрів $a = 2$ і $\gamma = 5$ розподілу Коші за вибіркою в 500 спостережень № 4

<table>
<tr><th>Спосіб №</th><th colspan="2">a^*</th><th>γ^*</th><th colspan="2">R^2</th></tr>
<tr><td>1</td><td colspan="2">0,596</td><td>2,771</td><td colspan="2">0,887</td></tr>
<tr><td>2</td><td colspan="2">0,363</td><td>6,106</td><td colspan="2">0,975</td></tr>
<tr><td>3</td><td colspan="2">1,485</td><td>5,736</td><td colspan="2">0,9983</td></tr>
<tr><td>4</td><td colspan="2">1,739</td><td>5,811</td><td colspan="2">**0,9992**</td></tr>
<tr><td>Квантильні оцінки</td><td>1,792</td><td>1,48</td><td>**5,558**</td><td>0,999</td><td>0,998</td></tr>
<tr><td>Оцінки ММВ</td><td colspan="2">**1,955**</td><td>5,718</td><td colspan="2">0,952</td></tr>
</table>

Таблиця 5. Результати моделювання параметрів $a = 2$ і $\gamma = 5$ розподілу Коші за всіма чотирма вибірками по 500 спостережень

Спосіб №	a^*		γ^*	R^2	
1	3,222		2,754	0,902	
2	2,995		6,148	0,984	
3	2,06		5,47	0,99956	
4	2,052		5,183	**0,9996**	
Квантильні оцінки	2,11	**1,992**	5,379	0,999	0,999
Оцінки ММВ	1,389		**5,13**	0,993	

Таблиця 6. Результати моделювання параметрів $a = 2$ і $\gamma = 5$ розподілу Коші за вибіркою в 50 спостережень

Спосіб №	a^*		γ^*	R^2	
1	-2,462		14,747	0,864	
2	-6,640		17,592	0,974	
3	0,783		7,796	0,9833	
4	1,486		**5,424**	**0,998**	
Квантильні оцінки	1,41	-0,88	7,17	0,982	0,925
Оцінки ММВ	**1,972**		6,022	0,996	

Таблиця 7. Результати моделювання параметрів $a = 2$ і $\gamma = 5$ розподілу Коші за вибіркою в 5000 спостережень

Спосіб №	a^*		γ^*	R^2	
1	1,64		5,464	0,943	
2	2,09		5,849	0,994	
3	1,782		6,609	0,9735	
4	2,02		**5,051**	**0,99998**	
Квантильні оцінки	2,01	**1,982**	5,105	0,9999	0,999
Оцінки ММВ	2,032		5,082	0,99996	

Зауваження 4. Наведемо для порівняння також оцінки параметрів розподілу Коші, отримані шляхом чисельного моделювання (у MathCad 15) для розглянутих вище вибірок за допомогою різних ітераційних процедур, запропонованих в [7, 9] та [10].

Згідно [7] для оцінки параметру масштабу γ розподілу Коші модифікованим методом моментів будемо використовувати моменти порядку $m = \dfrac{1}{k}$, де рекомендовані значення $4 \leq k \leq 5$.

Таблиця 8. Результати моделювання параметру масштабу $\gamma = 5$ розподілу Коші модифікованим методом моментів згідно [7, формула (4)] за розглянутими вище вибірками

	Вибірка об'єму			
Значення k	$n = 500$ № 1	$n = 500$ № 2	$n = 50$	$n = 5000$
3,0	5,277	6,105	7,015	9,391
4,0	5,329	6,058	6,779	9,843
4,1	5,332	6,055	6,76	9,876
4,2	5,334	6,052	6,741	9,908
4,3	5,337	6,05	6,724	9,939
4,4	5,339	6,047	6,707	9,968
4,5	5,341	6,045	6,691	9,996
4,6	5,343	6,043	6,676	10,022
4,7	5,345	6,041	6,661	10,048
4,8	5,347	6,039	6,646	10,072
4,9	5,348	6,037	6,632	10,096
5,0	5,35	6,036	6,619	10,119
6,0	5,361	6,024	6,507	10,305
10,0	5,375	6,008	6,276	10,687

Таблиця 9.1. Результати моделювання параметру $a = 2$ розподілу Коші згідно [9, формула (18)] за розглянутими вище вибірками

	Вибірка об'єму			
Крок k	$n = 500$ № 1	$n = 500$ № 2	$n = 50$	$n = 5000$
1	5,262	0,345	-0,246	0,312
2	2,039	0,933	-1,41	1,092
3	1,346	1,464	1,03	1,631
4	1,253	1,808	1,741	1,848
5	1,268	1,999	1,954	1,934
10	1,376	2,193	1,988	2,004
20	1,389	2,199	1,972	2,006

Таблиця 9.2. Результати моделювання параметру $\gamma = 5$ розподілу Коші згідно [9, формула (27)] за розглянутими вище вибірками

	Вибірка об'єму			
Крок k	$n = 500$ № 1	$n = 500$ № 2	$n = 50$	$n = 5000$
1	101,449	104,286	112,441	95,456
2	23,678	25,637	32,934	23,064
3	11,252	11,838	16,182	10,933
4	7,706	7,984	10,841	7,483
5	6,35	6,541	8,605	6,18
10	5,178	5,376	6,214	5,115
20	5,13	5,341	6,024	5,082

Таблиця 10.1. Результати моделювання параметру $a = 2$ розподілу Коші згідно [10] за розглянутими вище вибірками

	Вибірка об'єму			
Крок k	$n = 500$ № 1	$n = 500$ № 2	$n = 50$	$n = 5000$
30	5,384	10,152	2,56	11,506
35	2,843	4,862	1,609	4,319
40	1,958	3,251	1,265	2,714
45	1,61	2,65	1,133	2,269
50	1,467	2,405	1,08	2,112
100	1,366	2,231	1,043	2,005

Таблиця 10.2 Результати моделювання параметру $\gamma = 5$ розподілу Коші згідно [10] за розглянутими вище вибірками

	Вибірка об'єму			
Крок k	$n = 500$ № 1	$n = 500$ № 2	$n = 50$	$n = 5000$
30	8,543	13,339	8,001	63,62
35	6,035	7,203	7,295	17,541
40	5,425	5,818	7,174	7,868
45	5,284	5,497	7,17	5,757
50	5,254	5,419	7,181	5,266
100	5,249	5,405	7,194	5,123

4. Висновки

— незалежно від обсягу вибірки для отримання точніших оцінок варто відкидати частину вибірки згідно способів 3 або 4 пропозиції 3;

— зі збільшенням обсягу вибірки точність оцінок зростає;
— параметр γ запропонованим способом незалежно від обсягу вибірки оцінюється точніше, ніж квантильною оцінкою;
— оцінки за методом максимальної вірогідності мають середню точність, причому зі зростанням об'єму вибірки точність оцінок зростає;
— оцінки, утворені відкиданням частину вибірки згідно способів 3 або 4 пропозиції 3, мають приблизно однакову точність (проте є набагато простішими в обчисленні) порівняно з оцінками максимальної вірогідності;
— хоча, як правило, одна з двох квантильних оцінок параметру a є точнішою, проте наперед невідомо яка; запропонованим в пропозиції 3 способом параметр a оцінюється точніше, ніж середнє арифметичне двох квантильних його оцінок (0.1)–(0.2) незалежно від обсягу вибірки.

Лема 1. *При знаходженні оцінок параметрів за процедурою, як в пропозиції 3, тобто з відкиданням однакового числа найбільших і найменших вибіркових значень, оцінки невідомих теоретичних параметрів розподілу Коші (10)–(11) набувають більш зручного вигляду:*

$$a^* = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \tag{12}$$

$$\gamma^* = \frac{1}{A^*} = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum\limits_{i=1}^{n} x_i\right)^2}{\frac{1}{n}\sum\limits_{i=1}^{n} x_i z_i}. \tag{13}$$

Доведення. Формула (12) випливає з формул (7), (10) та з того, що при симетричному відкиданні найбільших і найменших вибіркових значень та при достатньо великому числі спостережень n середнє арифметичне $\bar{z} = \frac{1}{n}\sum\limits_{i=1}^{n} z_i = 0$.

Формула (13) випливає з аналогічних міркувань та з використанням формул (6), (11). □

Пропозиція 4. *Виходячи із властивостей арктангенса та природи соціально-економічних явищ, природно припустити, що арктангенсіальна залежність може описувати процес урбанізації (в країнах зі сталими соціально-економічними умовами); електоральні (виборчі) ті деякі інші соціально-економічні процеси; процес підвищення з* *часом ККД в різноманітних технологічних процесах (за умови відсутності технологічних революцій) тощо. Зрозуміло, що це стосується процесів зі сталими характеристиками розвитку (монотонність та необмеженість в часі) та з природно зрозумілими обмеженнями як знизу, так і зверху.*

5. Довірчі інтервали параметрів

Лема 2. *Для оцінки параметру розташування a при відомому параметрі масштабу γ довірчий інтервал надійності α має вигляд:*

$$\bar{x} - \gamma \tan\left(\frac{\pi}{2}\alpha\right) \le a \le \bar{x} + \gamma \tan\left(\frac{\pi}{2}\alpha\right).$$

Доведення. Результат леми 2 випливає з властивостей розподілу Коші. Як вже говорилось у вступі, величина $\bar{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$ теж буде розподіленою за законом розподілу Коші з тими ж параметрами. Тоді, за означенням

$$\mathbf{P}\{|\bar{x} - a| \le \delta\} = \frac{2}{\pi}\arctan\left(\frac{\delta}{\gamma}\right).$$

Зрозуміло, що при $\delta = \gamma \tan\left(\frac{\pi}{2}\alpha\right)$ отримаємо значення α в правій частині попередньої рівності, тобто шуканий довірчий інтервал надійності α. $\square$

Лема 3. *Для параметру масштабу γ при відомому параметрі розташування a довірчий інтервал надійності α має один з наступних виглядів:*

$$0 < \gamma \le \frac{|\bar{x} - a|}{\tan\left(\frac{\pi}{2}(1-\alpha)\right)};$$

$$\gamma \ge \frac{|\bar{x} - a|}{\tan\left(\frac{\pi}{2}\alpha\right)};$$

$$\frac{|\bar{x} - a|}{\tan\left(\frac{\pi}{4}(\alpha+1)\right)} \le \gamma \le |\bar{x} - a| \tan\left(\frac{\pi}{4}(\alpha+1)\right).$$

Доведення. Лема 3 доводиться аналогічно лемі 2. З тих же міркувань

1. $$\mathbf{P}\Big\{|\bar{x} - a| \ge \delta\gamma\Big\} = 1 - \frac{2}{\pi}\arctan\delta,$$

$$\mathbf{P}\Big\{\frac{|\bar{x} - a|}{\gamma} \ge \delta\Big\} = 1 - \frac{2}{\pi}\arctan\delta,$$

$$\mathbf{P}\Big\{\gamma \le \frac{|\bar{x}-a|}{\delta}\Big\} = 1 - \frac{2}{\pi}\arctan\delta,$$

і при $\delta = \tan\Big(\dfrac{\pi}{2}(1-\alpha)\Big)$ отримаємо значення α в правій частині попередньої рівності, тобто шуканий довірчий інтервал надійності α.

2. $$\mathbf{P}\Big\{|\bar{x}-a| \le \delta\gamma\Big\} = \frac{2}{\pi}\arctan\delta,$$

$$\mathbf{P}\Big\{\gamma \ge \frac{|\bar{x}-a|}{\delta}\Big\} = \frac{2}{\pi}\arctan\delta,$$

і при $\delta = \tan\Big(\dfrac{\pi}{2}\alpha\Big)$ отримаємо значення α в правій частині попередньої рівності, тобто шуканий довірчий інтервал надійності α.

3. $$\mathbf{P}\Big\{\frac{\gamma}{\delta} \le |\bar{x}-a| \le \delta\gamma\Big\} = \frac{2}{\pi}\Big(\arctan\delta - \arctan\frac{1}{\delta}\Big),$$

$$\mathbf{P}\Big\{\frac{1}{\delta} \le \frac{\gamma}{|\bar{x}-a|} \le \delta\Big\} = \frac{2}{\pi}\Big(\arctan\delta - \arctan\frac{1}{\delta}\Big),$$

$$\mathbf{P}\Big\{\frac{|\bar{x}-a|}{\delta} \le \gamma \le |\bar{x}-a|\delta\Big\} = \frac{2}{\pi}\Big(\arctan\delta - \arctan\frac{1}{\delta}\Big).$$

Використавши формулу

$$\arctan\frac{1}{\delta} = \frac{\pi}{2} - \operatorname{arccotan}\frac{1}{\delta} = \frac{\pi}{2} - \arctan\delta,$$

з останньої рівності будемо мати

$$\mathbf{P}\Big\{\frac{|\bar{x}-a|}{\delta} \le \gamma \le |\bar{x}-a|\delta\Big\} = \frac{4}{\pi}\arctan\delta - 1,$$

звідки при $\delta = \tan\Big(\dfrac{\pi}{4}(\alpha+1)\Big)$ отримаємо значення α в правій частині попередньої рівності, тобто шуканий довірчий інтервал надійності α.

$\square$

Подяка

Автор висловлює вдячність анонімному рецензенту за уважне вивчення тексту статті та корисні зауваження, які допомогли покращити роботу.

Додаток.
Графіки вибірки в 500 спостережень № 1 із розглянутих вище та змодельованих для неї ліній арктангенс-регресії

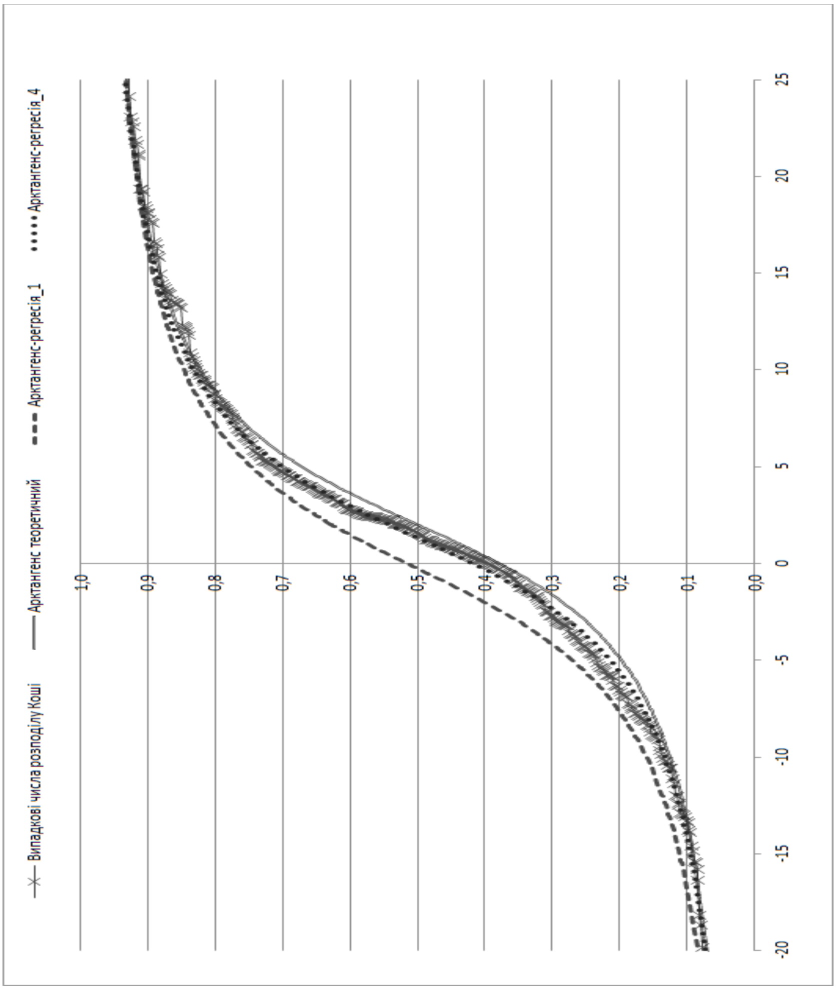

Рис. 1. Порівняння двох ліній арктангенс-регресії (за вибіркою, опрацьованою способами № 1 та № 4 Пропозиції 3) з вибірковими даними та з теоретичним розподілом Коші

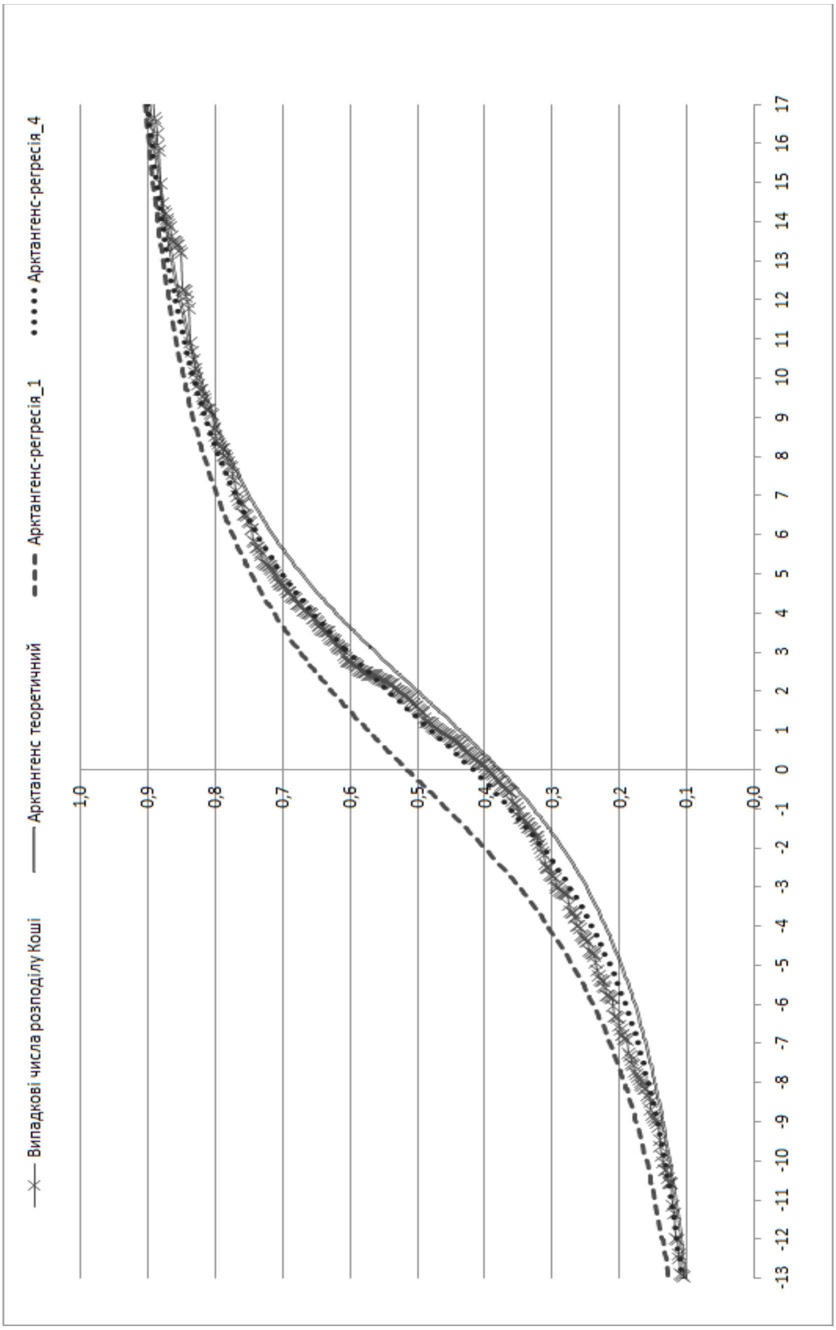

Рис. 2. Порівняння двох ліній арктангенс-регресії (за вибіркою, опрацьованою способами № 1 та № 4 Пропозиції 3) з вибірковими даними та з теоретичним розподілом Коші детальніше

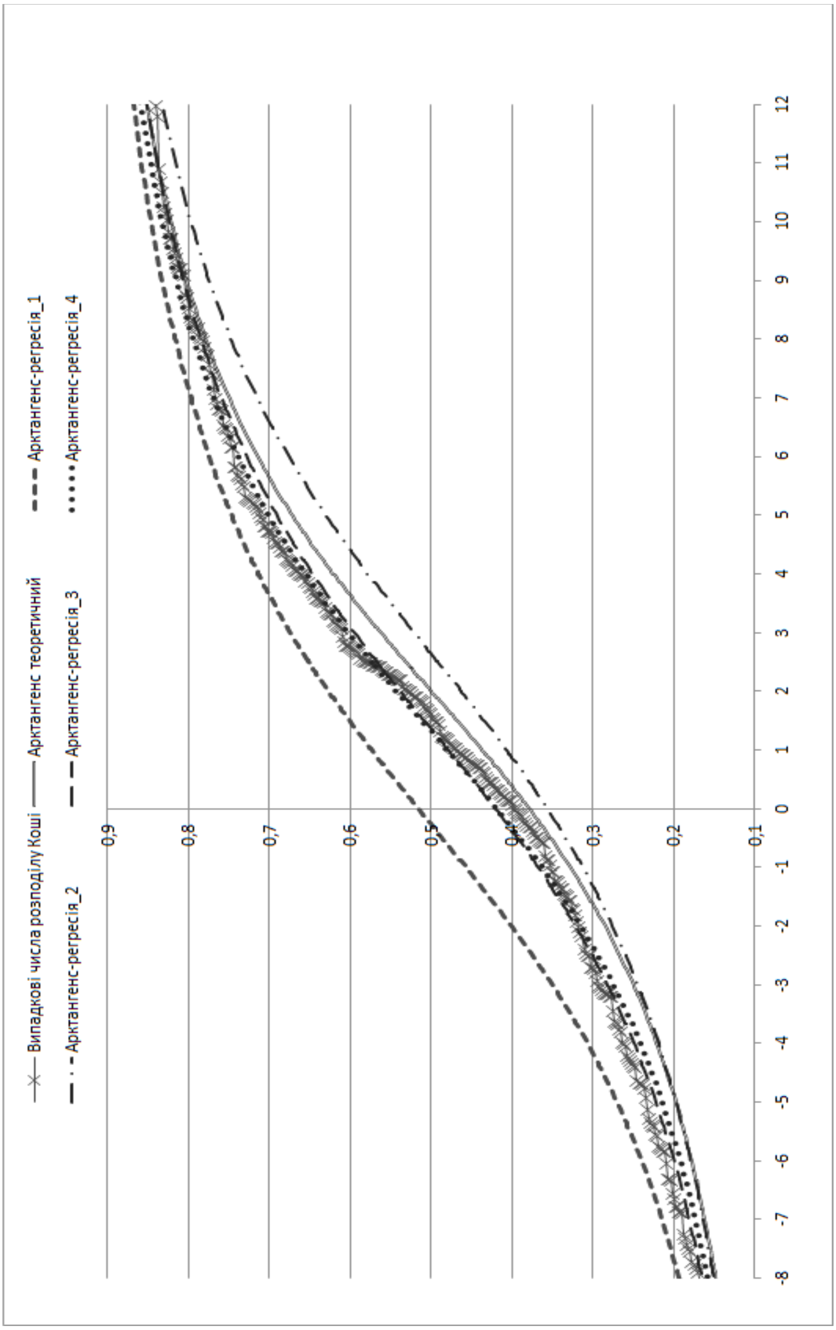

Рис. 3. Порівняння всіх чотирьох ліній арктангенс-регресії (за вибіркою, опрацьованою всіма способами Пропозиції 3) з вибірковими даними та з теоретичним розподілом Коші

Література

[1] Королюк, В.С., Портенко, Н.И., Скороход, А.В., Турбин, А.Ф. (1985). *Справочник по теории вероятностей и математической статистике.* Москва: Наука.

[2] Вадзинский, Р.Н. (2001). *Справочник по вероятностным распределениям.* Санктъ Петербургъ: Наука.

[3] Copas, J.B. (1975). On the unimodality of the likelihood for the Cauchy distribution. *Biometrika, 62*, 701–704.

[4] Ferguson, T.S. (1978). Maximum likelihood estimates of the parameters of the Cauchy distribution for samples of size 3 and 4. *Journal of the American Statistical Association, 73*, 211–213.

[5] Кендалл, М., Стьюарт, А. (1973). *Статистические выводы и связи, Т. 2.* Москва: Наука.

[6] Дубницкий, В.Ю., Ходырев, А.И. (2011). Оценивание параметров двупараметрического распределения Коши методом максимума правдоподобия для негруппированных и группированных наблюдений. *Системи обробки інформації, 2*(29), 17–20.

[7] Шинкеев, М.Л. (2012). Оценка параметров распределения Коши. *Научное обозрение, 3*, 77–81.

[8] Koutrouvelis, I.A. (1982). Estimation of location and scale in Cauchy distributions using the empirical characteristic function. *Biometrika, 65*(1), 205–213.

[9] Nagy, F. (2006). Parameter estimation of the Cauchy distribution in information theory approach. *Journal of Universal Computer Science, 12*, 1332–1344.

[10] Fegyverneki, S. (2013). A simple robust estimation for parameters of Cauchy distribution. *Miskolc Mathematical Notes, 14*(3), 887–892.

[11] Pekasiewicz, D. (2014). Application of quantile methods to estimation of Cauchy distribution parameters. *Statistics in transition, new series, 15*(1), 133–144.

[12] Галкин, В.М., Ерофеева, Л.Н., Лещева, С.В., Рыков, В.Е. (2015). Некоторые оценки параметров распределения Коши. *Труды НГТУ им. Алексеева, Нижний Новгород, 3*(110), 322–325.

[13] Sameer, A.H. (2016). Estimation of Cauchy Parameters under Ranked Set Sampling. *International Journal of Sciences: Basic and Applied Research, 28*(2), 284–295.

[14] Кузнецов, И.В., Писаренко, В.Ф., Родкин, М.В. (1997). Методы расчёта ущерба от катастроф различного типа. *Экономика и математические методы, 33*(4), 39–50.

[15] Абрамов, С.К. (2003). *Методи вторинної обробки сигналів та зображень у системах дистанційного зондування на основі використання міріадного оцінювання: автореф. дис. канд. техн. наук за фахом 05.07.12 – «Дистанційні аерокосмічні дослідження»*. Харків: Національний аерокосмічний університет ім. М.Є. Жуковського «ХАІ».

Відомості про авторів

Іван Григорович Крикун

Донецький національний університет
імені Василя Стуса,
Вінниця, Україна,
Інститут прикладної математики
і механіки НАН України,
Слов'янськ, Україна
E-Mail: `iwanko@i.ua`