Министерство высшего и среднего специального образования

Р С Ф С Р

Новосибирский электротехнический институт

УДК 519.25

В.И.Денисов, Б.Ю.Лемешко

ОЦЕНИВАНИЕ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ КОШИ ПО ЧАСТИЧНО ГРУППИРОВАННЫМ ВЫБОРКАМ

Новосибирск - 1979

ВВЕДЕНИЕ

Функция плотности распределения Коши имеет вид

$$f(x) = \frac{1}{\pi} \frac{\theta}{\theta^2 + (x-\theta_1)^2},$$

где θ - масштабный параметр, а θ_1 - параметр сдвига. Важные приложения этого распределения встречаются в физике [I] и теории надежности. Оно является примером паталогического распределения в статистике, поскольку его математическое ожидание не определено, а все другие моменты расходящиеся.

Пусть область определения случайной величины разбита на k интервалов групповыми пределами

$$-\infty < x_1 < x_2 < \ldots < x_{k-1} < +\infty.$$

Предполагается, что всегда $k \geq 2$. Пусть N - величина выборки, n_i - число наблюдений, попавших в i -ый интервал $(x_{i-1}, x_i]$, $\sum_{i=1}^{k} n_i = N$. Индивидуальные значения наблюдений, попадающих в i -ый интервал обозначим через $x_{i1}, x_{i2}, \ldots, x_{in_i}$.

Следуя [2], введем определение частично группированной выборки.

<u>Определение</u>. Выборка называется частично группированной, если имеющаяся в нашем распоряжении информация связана с множеством непересекающихся интервалов, которые делят область определения случайной величины так, что каждый интервал принадлежит к одному из 2-х типов :

а) i -ый интервал принадлежит к первому типу, если число n_i известно, но индивидуальные значения x_{ij} (j = I, 2, ... , ... , n_i) неизвестны ;

б) i -ый интервал принадлежит ко второму типу, если известно не только число n_i , но также и все индивидуальные значения x_{ij} (j = I, 2, ... , n_i).

Понятие частично группированной выборки объединяет в себе группированные, негруппированные и цензурированные выборки. Так мы имеем группированную выборку, если все интервалы принадлежат к первому типу, негруппированную, если все интервалы принадлежат ко второму типу, цензурированную слева (справа), если первый (k -ый) интервал принадлежит к первому типу, а все остальные ко второму, и имеет выборку с двусторонним цензурированием, если к первому типу принадлежат только первый и последний интервалы.

Суммирование и умножение по всем интервалам, принадлежащим к первому (или второму) типу, будем обозначать соответственно через $\sum\limits_{(1)}$ (или $\sum\limits_{(2)}$) и $\prod\limits_{(1)}$ (или $\prod\limits_{(2)}$). Интегрирование по всем интервалам второго типа обозначим $\int\limits_{(2)}$.

Для оценки неизвестного параметра по частично группированной выборке используется метод максимального правдоподобия. В нашем случае функция правдоподобия имеет вид

$$L = \prod_{(1)} p_i^{n_i} \prod_{(2)} \prod_{j=1}^{n_i} f(x_{ij}), \qquad (1)$$

где p_i - вероятность попадания наблюдения в i -ый интервал. Дифференцируя логарифм функции правдоподобия по неизвестному параметру θ и приравнивая его нулю, получаем уравнение правдоподобия

$$\sum_{(1)} n_i \frac{\partial \ln p_i}{\partial \theta} + \sum_{(2)} \sum_{j=1}^{n_i} \frac{\partial \ln f(x_{ij})}{\partial \theta} = 0. \qquad (2)$$

Оценка максимального правдоподобия (ОМП) неизвестного параметра получается в результате его решения. Как правило, это решение не может быть получено в явном виде и отыскивается с помощью различных итерационных методов.

К сожалению, ОМП неизвестного параметра по частично группированной выборке не всегда существует и единственна, что связано

с наличием интервалов первого типа. Это вызывает необходимость рассмотреть условия, при которых существуют и единственны оценки распределения Коши по частично группированной выборке и их асимптотические свойства.

ОЦЕНИВАНИЕ МАСШТАБНОГО ПАРАМЕТРА θ РАСПРЕДЕЛЕНИЯ КОШИ ПО ЧАСТИЧНО ГРУППИРОВАННОЙ ВЫБОРКЕ

Вероятность попадания наблюдения в i -ый интервал равна

$$p_i(\theta,\theta_1) = F(x_i;\theta,\theta_1) - F(x_{i-1};\theta,\theta_1) =$$

$$= \frac{1}{\pi}\,\mathrm{arctg}\, S_i - \frac{1}{\pi}\,\mathrm{arctg}\, S_{i-1},$$

где $S_i = (x_i-\theta_1)/\theta$. В дальнейшем вероятность попадания в интервал будем обозначать просто p_i .

<u>Теорема I</u>. ОМП масштабного параметра θ распределения Коши по частично группированной выборке существует тогда и только тогда, когда для интервалов первого типа выполняется одно из следующих условий :

- при $k \geq 2$, $n_1 < N$, $n_k < N$, $n_1 + n_k = N$ для $S_1 < S_{k-1} < 0$ $(x_1 < x_{k-1} < \theta_1)$

$$n_1 < n_k \frac{(1+S_1^2)\,S_{k-1}\,(\mathrm{arctg}\, S_1 + \pi/2)}{S_1(1+S_{k-1}^2)\,(\pi/2 - \mathrm{arctg}\, S_{k-1})},$$

для $0 < S_1 < S_{k-1}$ $(\theta_1 < x_1 < x_{k-1})$

$$n_1 > n_k \frac{(1+S_1^2)\,S_{k-1}\,(\mathrm{arctg}\, S_1 + \pi/2)}{S_1(1+S_{k-1})^2(\pi/2 - \mathrm{arctg}\, S_{k-1})};$$

- при $k > 2$, $n_1 + n_k < N$, для некоторого i такого, что $S_i < 0$ или $S_{i-1} > 0$ ($x_i < \theta_1$ или $x_{i-1} > \theta_1$), $n_i > 0$ или $\sum_{(2)} n_i > 0$. В этом случае ОМП получается в качестве корня уравнения правдоподобия

$$\frac{1}{\theta}\left\{\sum_{(1)} n_i \frac{\dfrac{S_{i-1}}{1+S_{i-1}^2} - \dfrac{S_i}{1+S_i^2}}{\operatorname{arctg} S_i - \operatorname{arctg} S_{i-1}} + \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^{n_i} \frac{2}{1+S_{ij}^2}\right\} = 0, \quad (3)$$

где $S_{ij} = (x_{ij} - \theta_1)/\theta$.

<u>Доказательство.</u> I. Покажем, что функции $\ln P_i$ и $\ln f(x)$ дважды дифференцируемы по θ :

$$\frac{\partial \ln P_i}{\partial \theta} = \frac{1}{\theta} \frac{\dfrac{S_{i-1}}{1+S_{i-1}^2} - \dfrac{S_i}{1+S_i^2}}{\operatorname{arctg} S_i - \operatorname{arctg} S_{i-1}},$$

$$\frac{\partial^2 \ln P_i}{\partial \theta^2} = \frac{1}{\theta^2}\left\{\frac{-2S_{i-1}}{(1+S_{i-1}^2)^2} + \frac{2S_i}{(1+S_i^2)^2} - \frac{\left(\dfrac{S_{i-1}}{1+S_{i-1}^2} - \dfrac{S_i}{1+S_i^2}\right)^2}{\operatorname{arctg} S_i - \operatorname{arctg} S_{i-1}}\right\} \Big/$$

$$(\operatorname{arctg} S_i - \operatorname{arctg} S_{i-1}), \quad (4)$$

$$\frac{\partial \ln f(x_{ij})}{\partial \theta} = \frac{1}{\theta}\left(1 - \frac{2}{1+S_{ij}^2}\right),$$

$$\frac{\partial^2 \ln f(x_{ij})}{\partial \theta^2} = \frac{1}{\theta^2}\left(2\frac{1-S_{ij}^2}{1+S_{ij}^2} - 1\right), \quad \text{для } i = 1, 2, \ldots, k.$$

2. Уравнение (3) получено из (2) подстановкой соответствующих значений производных. Так как $0 < \theta < +\infty$, то решение уравнения (3) эквивалентно решению уравнения

$$\sum_{(1)} n_i \frac{\dfrac{S_{i-1}}{1+S_{i-1}^2} - \dfrac{S_i}{1+S_i^2}}{\operatorname{arctg} S_i - \operatorname{arctg} S_{i-1}} + \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^{n_i} \frac{2}{1+S_{ij}^2} = 0. \quad (5)$$

Для доказательства существования решения уравнения (5), а, следовательно, и (3), необходимо показать, что найдутся такие значения параметра θ $\underline{\theta}$ и $\bar{\theta}$ ($\underline{\theta} < \bar{\theta}$), что при $\theta = \underline{\theta}$ левая часть уравнения (5) положительна, а при $\theta = \bar{\theta}$ - отрицательна.

Рассмотрим поведение левой части уравнения (5) при $\theta \to 0$. Применяя теорему Коши, имеем

$$\frac{\dfrac{s_{i-1}}{1+s_{i-1}^2} - \dfrac{s_i}{1+s_i^2}}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}} = \frac{\tau_i^2 - 1}{1+\tau_i^2}, \qquad (6)$$

где $s_{i-1} < \tau_i < s_i$ Заметим, что при $\theta \to 0$ $s_i^2 \to \infty$.

Для i -го интервала, принадлежащего первому типу, при $\theta \to 0$, $i = 2, \ldots, k-1$,

$$\lim_{\theta \to 0} \theta \frac{\partial \ln P_i}{\partial \theta} = \lim_{\tau_i^2 \to \infty} \frac{\tau_i^2 - 1}{\tau_i^2 + 1} = 1. \qquad (7)$$

Если $s_{i-1} < 0 < s_i$, то

$$\lim_{\theta \to 0} \theta \frac{\partial \ln P_i}{\partial \theta} = 0. \qquad (8)$$

Для $i = 1$ в случае $s_1 < 0$

$$\lim_{\theta \to 0} \theta \frac{\partial \ln P_1}{\partial \theta} = \lim_{s_1 \to -\infty} \left(- \frac{\dfrac{s_1}{1+s_1^2}}{\operatorname{arctg} s_1 + \pi/2} \right) = 1, \qquad (9)$$

при $s_1 > 0$

$$\lim_{\theta \to 0} \theta \frac{\partial \ln P_1}{\partial \theta} = 0. \qquad (10)$$

Для $i = k$ при $s_{k-1} > 0$

$$\lim_{\theta \to 0} \theta \frac{\partial \ln P_k}{\partial \theta} = \lim_{s_{k-1} \to +\infty} \frac{\dfrac{s_{k-1}}{1+s_{k-1}^2}}{\pi/2 - \operatorname{arctg} s_{k-1}} = 1, \qquad (11)$$

при $s_{k-1} < 0$

$$\lim_{\theta \to 0} \theta \frac{\partial \ln P_k}{\partial \theta} = \lim_{s_{k-1} \to -\infty} \frac{\dfrac{s_{k-1}}{1+s_{k-1}^2}}{\pi/2 - \operatorname{arctg} s_{k-1}} = 0. \qquad (12)$$

Для всех интервалов второго типа

$$\lim_{s_{ij}^2 \to \infty} \left(\sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^{n_i} \frac{2}{1+s_{ij}^2} \right) \doteq \sum_{(2)} n_i. \qquad (13)$$

Из (7-13) очевидно, что при условии $n_1 < N$ при $S_1 > 0$, или $n_k < N$ при $S_{k-1} < 0$, или $n_i < N$ при $S_{i-1} < 0 < S_i$ можно указать некоторое $\underline{\theta}$, для которого выполняется неравенство

$$\left(\sum_{(1)} n_i \frac{\frac{S_{i-1}}{1+S_{i-1}^2} - \frac{S_i}{1+S_i^2}}{\operatorname{arctg} S_i - \operatorname{arctg} S_{i-1}} + \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^{n_i} \frac{2}{1+S_{ij}^2} \right) \Bigg|_{\theta=\underline{\theta}} > 0. \quad (14)$$

Рассмотрим, как меняется левая часть уравнения (5) при $\theta \to +\infty$. Для i-го интервала, $i = 2, \ldots, k-1$, принадлежащего первому типу, имеем

$$\lim_{\theta \to \infty} \theta \frac{\partial \ln P_i}{\partial \theta} = \lim_{\tau_i^2 \to 0} \frac{\tau_i^2 - 1}{\tau_i^2 + 1} = -1. \quad (15)$$

Для $i = 1$ и $S_1 < 0$

$$\lim_{\theta \to \infty} \theta \frac{\partial \ln P_1}{\partial \theta} = \lim_{S_1 \to 0} \left(- \frac{\frac{S_1}{1+S_1^2}}{\operatorname{arctg} S_1 + \pi/2} \right) = 0_+ . \quad (16)$$

при $S_1 > 0$

$$\lim_{\theta \to \infty} \theta \frac{\partial \ln P_1}{\partial \theta} = 0_- . \quad (17)$$

Для $i = k$ и $S_{k-1} > 0$

$$\lim_{\theta \to \infty} \theta \frac{\partial \ln P_k}{\partial \theta} = \lim_{S_{k-1} \to 0} \frac{\frac{S_{k-1}}{1+S_{k-1}^2}}{\pi/2 - \operatorname{arctg} S_{k-1}} = 0_+ . \quad (18)$$

при $S_{k-1} < 0$

$$\lim_{\theta \to \infty} \theta \frac{\partial \ln P_k}{\partial \theta} = 0_- . \quad (19)$$

Для всех интервалов второго типа

$$\lim_{S_{ij}^2 \to 0} \left(\sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^{n_i} \frac{2}{1+S_{ij}^2} \right) = - \sum_{(2)} n_i . \quad (20)$$

Необходимо отдельно рассмотреть случай, когда $n_1 + n_k = N$.

Рассмотрим неравенство

$$-n_1 \frac{\frac{s_1}{1+s_1^2}}{\operatorname{arctg} s_1 + \pi/2} + n_k \frac{\frac{s_{k-1}}{1+s_{k-1}}}{\pi/2 - \operatorname{arctg} s_{k-1}} < 0. \qquad (21)$$

Это неравенство выполняется, если

$$n_1 > n_k \frac{(1+s_1^2)\, s_{k-1} (\operatorname{arctg} s_1 + \pi/2)}{s_1 (1+s_{k-1}^2)(\pi/2 - \operatorname{arctg} s_{k-1})} \qquad (22)$$

для $0 < s_1 < s_{k-1}$ и

$$n_1 < n_k \frac{(1+s_1^2)\, s_{k-1} (\operatorname{arctg} s_1 + \pi/2)}{s_1 (1+s_{k-1}^2)(\pi/2 - \operatorname{arctg} s_{k-1})} \qquad (23)$$

для $s_1 < s_{k-1} < 0$.

Из (15-21) следует, что при $n_1 < N$ для $s_1 < 0$, $n_k < N$ для $s_{k-1} > 0$ и $n_1 + n_k < N$ для $s_1 \leq 0 \leq s_{k-1}$ или $n_1 + n_k = N$ при выполнении условий (22) или (23) соответственно, можно указать некоторое $\bar{\theta}$, для которого выполняется неравенство

$$\left(\sum_{(1)} n_i \frac{\frac{s_{i-1}}{1+s_{i-1}^2} - \frac{s_i}{1+s_i^2}}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}} + \sum_{(2)} n_i - \sum_{(2)} \sum_{j=1}^{n_i} \frac{2}{1+s_{ij}^2} \right) \Bigg|_{\theta=\bar{\theta}} < 0. \qquad (24)$$

Из (14) и (24) следует существование решения уравнения правдоподобия (3) в условиях теоремы.

Для полностью группированной выборки условия существования ОМП параметра θ определяются следующей теоремой.

<u>Теорема 2.</u> ОМП масштабного параметра θ распределения Коши по группированной выборке существует тогда и только тогда, когда: при $k = 2$, $n_1 < N$, $n_2 < N$ для $s_1 < 0$ $(x_1 < \theta_1)$

$$n_1 < n_2 \frac{\operatorname{arctg} s_1 + \pi/2}{\pi/2 - \operatorname{arctg} s_1};$$

для $s_1 > 0$ $(x_1 > \theta_1)$

$$n_1 > n_2 \frac{\operatorname{arctg} s_1 + \pi/2}{\pi/2 - \operatorname{arctg} s_1};$$

при $k>2$, $n_1<N$, $n_k<N$, $n_1+n_k=N$ для $s_1<s_{k-1}<0$ ($x_1<x_{k-1}<\theta_1$)

$$n_1 < n_k \frac{(1+s_1^2)s_{k-1}(\operatorname{arctg} s_1 + \pi/2)}{s_1(1+s_{k-1}^2)(\pi/2 - \operatorname{arctg} s_{k-1})},$$

для $0<s_1<s_{k-1}$ ($\theta_1<x_1<x_{k-1}$)

$$n_1 > n_k \frac{(1+s_1^2)s_{k-1}(\operatorname{arctg} s_1 + \pi/2)}{s_1(1+s_{k-1}^2)(\pi/2 - \operatorname{arctg} s_{k-1})};$$

при $k>2$, $n_1+n_k<N$ и для некоторого i такого, что $s_i<0$ ($x_i<\theta_1$) или $s_{i-1}>0$ ($x_{i-1}>\theta_1$), $n_i>0$. В этом случае ОМП параметра θ определяется как решение уравнения правдоподобия

$$\frac{1}{\theta}\sum_{i=1}^{k} n_i \frac{\dfrac{s_{i-1}}{1+s_{i-1}^2} - \dfrac{s_i}{1+s_i^2}}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}} = 0. \qquad (25)$$

В частном случае при $k=2$ для ОМП параметра θ можно найти аналитическое выражение :

$$\hat{\theta} = (x_1 - \theta_1)\operatorname{ctg}\left(\frac{\pi}{2}\,\frac{2n_1 - N}{N}\right).$$

<u>Теорема 3.</u> ОМП масштабного параметра θ распределения Коши по группированным данным состоятельна и асимптотически эффективна, если она существует и единственна.

<u>Доказательство.</u> Первые частные производные функций P_i по θ имеют вид

$$\frac{\partial P_i}{\partial \theta} = \frac{1}{\theta\pi}\left(\frac{s_{i-1}}{1+s_{i-1}^2} - \frac{s_i}{1+s_i^2}\right);$$

и, следовательно, так как $k\geq 2$, не все $\partial P_i/\partial\theta$ равны нулю

при истинном значении параметра θ. Вторые частные производные, определяемые выражением

$$\frac{\partial^2 p_i}{\partial \theta^2} = \frac{1}{\theta^2 \pi}\left[\frac{-2\cdot s_{i-1}}{(1+s_{i-1}^2)^2} + \frac{2\cdot s_i}{(1+s_i^2)^2}\right],$$

также непрерывны. Справедливость теоремы следует из теорем 3 и 5 § I [3].

Асимптотическая дисперсия ОМП масштабного параметра θ по группированным данным равна

$$D(\hat{\theta}) = N^{-1} J_r^{-1}(\hat{\theta}),$$

где $J_r(\theta)$ - информационное количество Фишера о параметре θ на одно наблюдение по группированным данным и находится из соотношения

$$J_r(\theta) = \sum_{i=1}^{k}\left(\frac{\partial \ln p_i}{\partial \theta}\right)^2 p_i = \frac{1}{\theta^2 \pi}\sum_{i=1}^{k}\frac{\left(\frac{s_{i-1}}{1+s_{i-1}^2} - \frac{s_i}{1+s_i^2}\right)^2}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}}. \quad (26)$$

Зависимость $J_r(\theta)$ от групповых пределов позволяет подобрать их таким образом, чтобы минимизировать асимптотическую дисперсию или, что то же самое, максимизировать $J_r(\theta)$. В таблице I представлены вычисленные асимптотически оптимальные групповые пределы в виде $s_i = (x_i - \theta_1)/\theta$. В последнем столбце таблицы приведены значения относительной асимптотической эффективности A по сравнению с оцениванием по негруппированной :

$$A = \frac{J_r(\theta)}{J_н(\theta)},$$

где $J_н(\theta)$ - количество информации Фишера на одно наблюдение по негруппированным данным

$$J_н(\theta) = \int_{-\infty}^{\infty}\left(\frac{\partial \ln f(x)}{\partial \theta}\right)^2 f(x)\,dx = \frac{1}{2\theta^2}.$$

Таблица I. Оптимальные групповые пределы для оценивания масштабного параметра θ распределения Коши и проверки гипотез о нём по критерию χ^2 Пирсона.

k	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	A
2	-I.3274										0.2867
	I.3274										0.2867
3	-I.0000	I.0000									0.8I06
4	-0.9597	0.7I49	I.5797								0.8630
	-I.5797	-0.7I49	0.9597								0.8630
5	-I.5I7I	-0.659I	0.659I	I.5I7I							0.9226
6	-I.4959	-0.6398	0.5438	I.0228	I.9562						0.9339
	-I.9562	-I.0228	-0.5438	0.6398	I.4959						0.9339
7	-I.9269	-I.0000	-0.5I90	0.5I90	I.0000	I.9269					0.9582
8	-I.9I30	-0.989I	-0.5070	0.4529	0.80I5	I.2806	2.2993				0.97I5
	-2.2993	-I.2806	-0.80I5	-0.4529	0.5070	0.989I	I.9I30				0.97I5
9	-2.28I4	-I.2674	-0.789I	-0.4384	0.4384	0.789I	I.2674	2.28I4			0.9739
I0	-2.27I2	-I.2600	-0.78I9	-0.4300	0.3944	0.6759	I.0078	I.5060	2.6I24		0.9780
	-2.6I24	-I.5060	-I.0078	-0.6579	-0.3944	0.4300	0.78I9	I.2600	2.27I2		0.9780
II	-2.6002	-I.4972	-I.0000	-0.6679	-0.3846	0.3846	0.6679	I.0000	I.4972	2.6002	0.982I

Непосредственное использование таблицы I для построения оценок масштабного параметра θ по группированным данным, возможно, благодаря априорным сведениям об этом параметре, по которым и строятся групповые пределы.

Но особенное значение асимптотически оптимальные групповые пределы имеют при проверке гипотезы о параметре θ по критерию χ^2 Пирсона. Применение асимптотически оптимальных групповых пределов позволяют минимизировать потери информации, вызываемые группированием исходной выборки, и, следовательно, увеличить мощность критерия χ^2 при близких альтернативных гипотезах. Так потери в информационном количестве Фишера при k = II составляют менее 2% , что значительно ниже потерь при равномерном или равновероятном группировании, наиболее употребляемых на практике.

ОЦЕНИВАНИЕ ПАРАМЕТРА СДВИГА θ_1 РАСПРЕДЕЛЕНИЯ КОШИ ПО ЧАСТИЧНО ГРУППИРОВАННОЙ ВЫБОРКЕ

<u>Теорема 4.</u> ОМП параметра сдвига θ_1 распределения Коши по частично группированной выборке существует тогда и только тогда, когда $\sum_{(2)} n_i > 0$ или для интервалов первого типа $n_1 < N$ и $n_k < N$. При этом ОМП получается в качестве решения уравнения правдоподобия

$$\frac{1}{\theta}\left\{\sum_{(1)} n_i \frac{\frac{1}{1+S_{i-1}^2} - \frac{1}{1+S_i^2}}{\operatorname{arctg} S_i - \operatorname{arctg} S_{i-1}} + 2\sum_{(2)}\sum_{j=1}^{n_i} \frac{S_{ij}}{1+S_{ij}^2}\right\} = 0. \quad (27)$$

<u>Доказательство</u>: I. Покажем, что функции $\ln p_i$ и $\ln f$ дважды дифференцируемы по θ_1 :

$$\frac{\partial \ln p_i}{\partial \theta_1} = \frac{1}{\theta}\,\frac{\frac{1}{1+S_{i-1}^2} - \frac{1}{1+S_i^2}}{\operatorname{arctg} S_i - \operatorname{arctg} S_{i-1}} \; ? \; ;$$

$$\frac{\partial^2 \ln p_i}{\partial \theta_1^2} = \frac{1}{\theta^2}\left\{\frac{2 s_{i-1}}{(1+s_{i-1}^2)^2} - \frac{2\cdot s_i}{(1+s_i^2)^2} - \frac{\left(\frac{1}{1+s_{i-1}^2} - \frac{1}{1+s_i^2}\right)^2}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}}\right\} /$$

$$(\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}), \qquad (28)$$

$$\frac{\partial \ln f}{\theta \theta_1} = \frac{1}{\theta}\,\frac{2 s_{ij}}{1+s_{ij}^2},$$

$$\frac{\partial^2 \ln f}{\partial \theta_1^2} = \frac{2}{\theta^2}\,\frac{s_{ij}^2 - 1}{(1+s_{ij}^2)^2}.$$

2. Уравнение (27) получено из (2) подстановкой соответствующих $\partial \ln p_i / \partial \theta_1$ и $\partial \ln f / \partial \theta_1$. Так как $0 < \theta < \infty$, то будем рассматривать уравнение

$$\sum_{(1)} n_i \frac{\frac{1}{1+s_{i-1}^2} - \frac{1}{1+s_i^2}}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}} + 2 \sum_{(2)} \sum_{j=1}^{n_i} \frac{s_{ij}}{1+s_{ij}^2} = 0, \qquad (29)$$

имеющее те же корни, что и (27).

Для доказательства существования ОМП необходимо показать, что существуют некоторые $\underline{\theta}_1$ и $\bar{\theta}_1$, $\underline{\theta}_1 < \bar{\theta}_1$, такие, что левая часть равенства (29) положительна при $\theta_1 = \underline{\theta}_1$ и отрицательна при $\theta_1 = \bar{\theta}_1$.

Рассмотрим поведение левой части уравнения (29) при $\theta_1 \to -\infty$. Применяя теорему Коши, имеем

$$\frac{\frac{1}{1+s_{i-1}^2} - \frac{1}{1+s_i^2}}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}} = \frac{2\tau_i}{1+\tau_i^2}, \qquad (30)$$

где $s_{i-1} < \tau_i < s_i$.

Для i-го интервала первого типа, $i = 2, \ldots, k$, используя (30), получаем

$$\lim_{\theta_1 \to -\infty} \theta \frac{\partial \ln p_i}{\partial \theta_1} = \lim_{\tau_i \to +\infty} \frac{2\tau_i}{1+\tau_i^2} = 0_+. \qquad (31)$$

Для $i = I$

$$\lim_{\theta_1 \to -\infty} \theta \frac{\partial \ln P_1}{\partial \theta_1} = \lim_{s_1 \to +\infty} \left\{ - \frac{\frac{1}{1+s_1^2}}{\operatorname{arctg} s_1 + \pi/2} \right\} = 0_- . \qquad (32)$$

Для всех интервалов второго типа имеем

$$\lim_{\theta_1 \to -\infty} \theta \sum_{(2)} \sum_{j=1}^{n_i} \frac{\partial \ln f(x_{ij})}{\partial \theta_1} = \lim_{s_{ij} \to +\infty} 2 \sum_{(2)} \sum_{j=1}^{n_i} \frac{s_{ij}}{1+s_{ij}^2} = 0_+ . \qquad (33)$$

Из (3I-33) очевидно, что при условии $n_1 < N$, если I-ый интервал относится к первому типу, найдется некоторое $\underline{\theta}_1$, для которого выполняется неравенство

$$\left\{ \sum_{(1)} n_i \frac{\frac{1}{1+s_{i-1}^2} - \frac{1}{1+s_i^2}}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}} + 2 \sum_{(2)} \sum_{j=1}^{n_i} \frac{s_{ij}}{1+s_{ij}^2} \right\} \Bigg|_{\theta_1 = \underline{\theta}_1} > 0. \qquad (34)$$

Далее, при $\theta_1 \to +\infty$ для $i = I, \ldots, k - I$, имеем

$$\lim_{\theta_1 \to +\infty} \theta \frac{\partial \ln P_i}{\partial \theta_1} = \lim_{\tau_i \to -\infty} \frac{2\tau_i}{1+\tau_i^2} = 0_- . \qquad (36)$$

Для $i = k$

$$\lim_{\theta_1 \to +\infty} \theta \frac{\partial \ln P_k}{\partial \theta_1} = \lim_{s_{k-1} \to -\infty} \frac{-\frac{1}{1+s_{k-1}^2}}{\pi/2 - \operatorname{arctg} s_{k-1}} = 0_+ . \qquad (37)$$

Для всех интервалов второго типа

$$\lim_{\theta_1 \to +\infty} \theta \sum_{(2)} \sum_{j=1}^{k} \frac{\partial \ln f(x_{ij})}{\partial \theta_1} = \lim_{s_{ij} \to -\infty} 2 \sum_{(2)} \sum_{j=1}^{n_i} \frac{s_{ij}}{1+s_{ij}^2} = 0_- . \qquad (38)$$

Из (36) - (38) очевидно, что при условии $n_k < N$ для интервала первого типа найдется некоторое $\bar{\theta}_1$ такое, для которого справедливо неравенство

$$\left\{ \sum_{(1)} n_i \frac{\frac{1}{1+s_{i-1}^2} - \frac{1}{1+s_i^2}}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}} + 2 \sum_{(2)} \sum_{j=1}^{n_i} \frac{s_{ij}}{1+s_{ij}^2} \right\} \Bigg|_{\theta_1 = \bar{\theta}_1} < 0. \qquad (39)$$

Из (34) и (39) следует существование ОМП параметра θ_1 в условиях теоремы.

В частном случае для группированной выборки справедлива следующая теорема.

Теорема 5. ОМП параметра сдвига θ_1 распределения Коши существует тогда и только тогда, когда $n_1 < N$ и $n_k < N$ и получается в качестве решения уравнения правдоподобия

$$\frac{1}{\theta}\sum_{i=1}^{k} n_i \frac{\frac{1}{1+s_{i-1}^2} - \frac{1}{1+s_i^2}}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}} = 0 . \qquad (40)$$

При $k = 2$ ОМП параметра θ_1 определяется выражением

$$\hat{\theta}_1 = x_1 - \theta \cdot tg\left(\frac{\pi}{2}\,\frac{2n_1 - N}{N}\right).$$

Теорема 6. ОМП параметра сдвига θ_1 распределения Коши по группированным данным состоятельна и асимптотически эффективна, если она существует и единственна.

Доказательство аналогично доказательству теоремы 3 .

Информационное количество Фишера о параметре θ_1 на одно наблюдение по группированным данным описывается соотношением

$$J_r(\theta_1) = \sum_{i=1}^{k}\left(\frac{\partial \ln p_i}{\partial \theta_1}\right)^2 p_i = \frac{1}{\theta^2 \pi}\sum_{i=1}^{k}\frac{\left(\frac{1}{1+s_{i-1}^2} - \frac{1}{1+s_i^2}\right)^2}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}} . \qquad (41)$$

Асимптотически оптимальные групповые пределы в виде $s_i = (x_i - \theta_1)/\theta$, максимизирующие информационное количество, а, следовательно, минимизирующие асимптотическую дисперсию ОМП $\hat{\theta}_1$, и соответствующие значения относительной асимптотической эффективности

$$A = J_r(\theta_1) / J_H(\theta_1), \qquad \text{где}$$

$$J_H(\theta_1) = \int_{-\infty}^{\infty}\left(\frac{\partial \ln f(x)}{\partial \theta_1}\right)^2 f(x)\,dx = \frac{1}{2\theta^2}$$

- информационное количество Фишера о параметре θ_1, по негруппи-

Таблица 2. Оптимальные групповые пределы для оценивания параметра сдвига θ распределения Коши и проверки гипотез о нём по критерию χ^2 Пирсона.

k	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	A
2	0.0											0.8106
3	-0.1794	0.1794										0.8577
4	-3.2845	-0.2390	0.1570									0.8991
	-0.1570	0.2390	3.2845									0.8991
5	-3.3338	-0.2137	0.2137	3.3338								0.9393
6	-3.1577	-0.3168	0.0	0.3168	3.1577							0.9582
7	-3.0827	-0.3786	-0.1150	0.1150	0.3786	3.0827						0.9658
8	-5.0956	-2.2911	-0.4018	-0.1275	0.1072	0.3744	3.0863					0.9718
	-3.0863	-0.3744	-0.1072	0.1275	0.4018	2.2911	5.0956					0.9718
9	-5.1041	-2.2961	-0.3972	-0.1196	0.1196	0.3972	2.2961	5.1041				0.9779
10	-5.0213	-2.2498	-0.4445	-0.1991	0.0	0.1991	0.4445	2.2498	5.0213			0.9821
11	-4.9723	-2.2216	-0.4792	-0.2572	-0.0820	0.0820	0.2572	0.4792	2.2216	4.9723		0.9845
12	-4.9357	-2.2027	-0.5059	-0.3020	-0.1433	0.0	0.1433	0.3020	0.5059	2.2027	4.9357	0.9860

рованным данным, представлены в таблице 2. Последняя может использоваться при оценивании параметра θ_1 по группированным данным, причем оптимальные групповые пределы будут определяться по прогнозируемому значению параметра θ_1, а также при проверке гипотеза о параметре θ_1 по критерию χ^2.

ОДНОВРЕМЕННОЕ ОЦЕНИВАНИЕ ДВУХ ПАРАМЕТРОВ РАСПРЕДЕЛЕНИЯ КОШИ

Существование ОМП вектора параметров распределения Коши по группированным данным при $k>2$ определяется непосредственно условиями теорем 2 и 5 на основании теоремы 2 § I [3]. ОМП получается в качестве решения системы уравнений (25) и (39). Если ОМП по группированным данным существует и единственна, то она будет состоятельна и асимптотически эффективна. Асимптотическая дисперсионная матрица ОМП параметров распределения Коши по группированным данным равна

$$D(\hat{\theta},\hat{\theta}_1) = N^{-1} J_r^{-1}(\hat{\theta},\hat{\theta}_1),$$

где $J_r(\theta,\theta_1)$ - информационная матрица Фишера по группированным данным, равная

$$J_r(\theta,\theta_1) = \begin{bmatrix} \sum_{i=1}^{k} p_i \left(\frac{\partial \ln p_i}{\partial \theta}\right)^2 & \sum_{i=1}^{k} p_i \frac{\partial \ln p_i}{\partial \theta}\frac{\partial \ln p_i}{\partial \theta_1} \\ \sum_{i=1}^{k} p_i \frac{\partial \ln p_i}{\partial \theta_1}\frac{\partial \ln p_i}{\partial \theta} & \sum_{i=1}^{k} p_i \left(\frac{\partial \ln p_i}{\partial \theta_1}\right)^2 \end{bmatrix}.$$

Элементы главной диагонали матрицы $J_r(\theta,\theta_1)$ описываются соотношениями (26) и (41), а

$$\sum_{i=1}^{k} p_i \frac{\partial \ln p_i}{\partial \theta}\frac{\partial \ln p_i}{\partial \theta_1} = \frac{1}{\theta^2 \pi} \sum_{i=1}^{k} \frac{\left(\frac{s_{i-1}}{1+s_{i-1}^2} - \frac{s_i}{1+s_i^2}\right)\left(\frac{1}{1+s_i^2} - \frac{1}{1+s_{i-1}^2}\right)}{\operatorname{arctg} s_i - \operatorname{arctg} s_{i-1}}.$$

Таблица 3. Асимптотически оптимальные групповые пределы для совместного оценивания параметров θ и θ_1 распределения Коши по группированным данным и проверки согласия по критерию χ^2 Пирсона.

k	S_1	S_2	S_3	S_4	S_5	S_6	S_7
3	-0.5773	0.5773					
4	-I.0000	0.0	I.0000				
5	-I.3764	-0.3249	0.3249	I.3764			
6	-I.7320	-0.5773	0.0	0.5773	I.7320		
7	-2.0765	-0.7975	-0.2282	0.2282	0.7975	2.0765	
8	-2.4I42	-I.0000	-0.4I42	0.0	0.4I42	I.0000	2.4I42
9	-2.7475	-I.I9I7	-0.5773	-0.I763	0.I763	0.5773	I.I9I7
I0	-3.0778	-I.3764	-0.7266	-0.3249	0.0	0.3249	0.7266
II	-3.4062	-I.556I	-0.8665	-0.4567	-0.I438	0.I438	0.4567
I2	-3.73I9	-I.7320	-I.0000	-0.5773	-0.2679	0.0	0.2679
I3	-4.0573	-I.9053	-I.I288	-0.6903	-0.3793	-0.I2I4	0.I2I4
I4	-4.3809	-2.0763	-I.2538	-0.7974	-0.48I5	-0.2282	0.0
I5	-4.704I	-2.2467	-I.3762	-0.9003	-0.5773	-0.3249	-0.I05I

k	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	A
3								0.4677
4								0.6570
5								0.7659
6								0.83I5
7								0.8735
8								0.90I8
9	2.7475							0.92I7
I0	I.3764	3.0778						0.936I
II	0.8665	I.556I	3.4062					0.9469
I2	0.5773	I.0000	I.7320	3.73I9				0.9552
I3	0.3793	0.6903	I.I288	I.9053	4.0573			0.96I7
I4	0.2282	0.48I5	0.7974	I.2538	2.0763	4.3809		0.9669
I5	0.I05I	0.3249	0.5773	0.9003	I.3762	2.2457	4.704I	0.97II

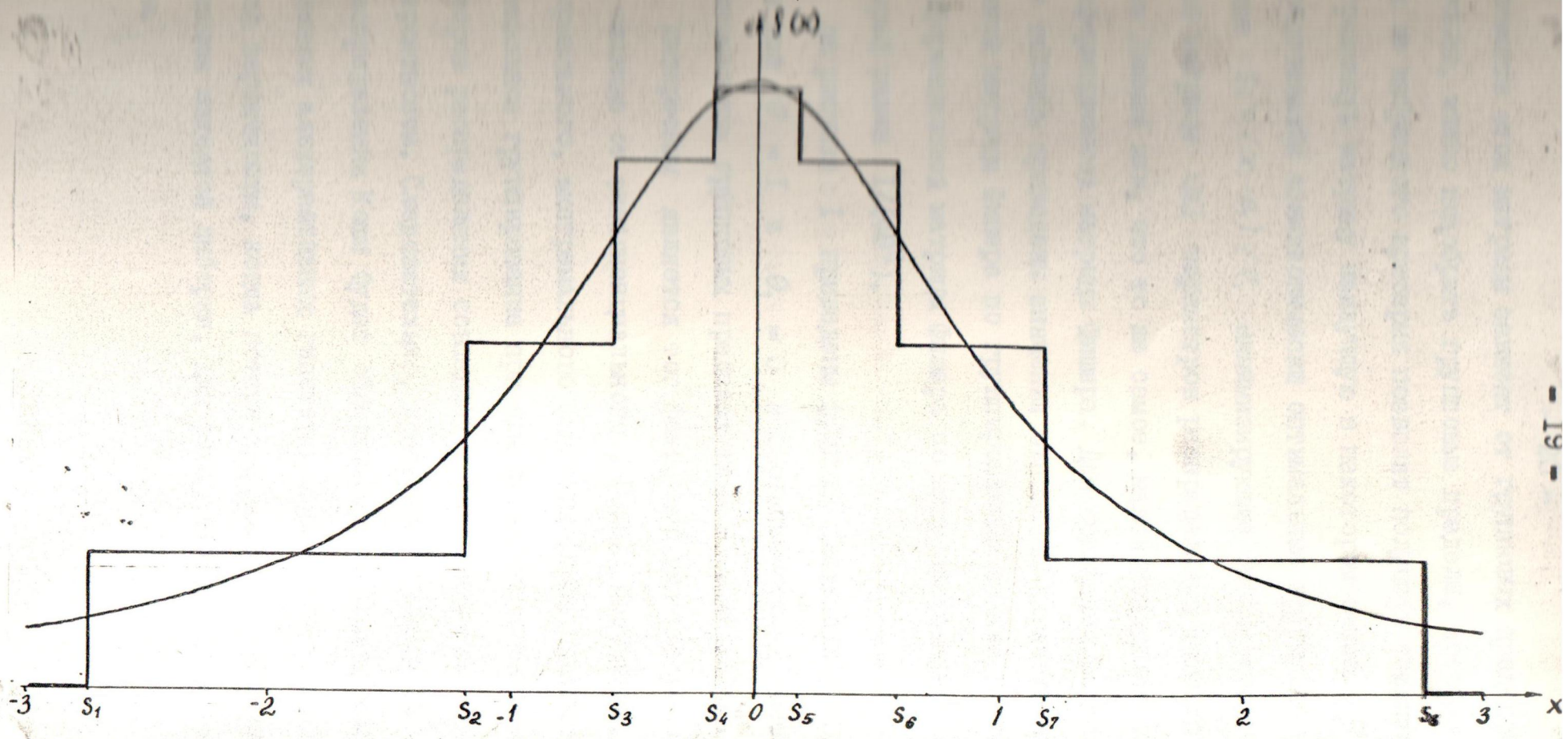

Рис. I. Функция плотности распределения Коши при θ = I, θ_1 = 0 и соответствующая гистограмма при оптимальном разбиении области определения случайной величины на 9 интервалов.

Элементы этой матрицы зависят от групповых пределов. Следовательно, можно подобрать групповые пределы, которые в зависимости от выбранного критерия позволят получить асимптотическую дисперсионную матрицу наилучшую в некотором смысле. В таблице 3 представлены асимптотически оптимальные групповые пределы в виде $S_i = (x_i - \theta_1)/\theta$, минимизирующие определитель дисперсионной матрицы ОМП параметров распределения Коши по группированным данным или, что то же самое, максимизирующие определитель информационной матрицы Фишера. Для сравнения в последнем столбце таблицы приведены значения отношения определителя информационной матрицы Фишера по группированным данным к определителю информационной матрицы Фишера по негруппированной выборке, который равен $I/(4\theta^4)$.

На рисунке I приведены функция плотности распределения Коши при $\theta = I$ и $\theta_1 = 0$ и соответствующая гистограмма при оптимальных групповых пределах для 9 интервалов.

Интересным является тот факт, что для распределения Коши, в отличие от экспоненциального, Рэлея, Максвелла, Вейбулла, нормального, экстремального значения [4-5], асимптотически оптимальное группирование при одновременном оценивании двух параметров распределения совпадает с разбиением на интервалы равной вероятности. Следовательно, критерий согласия χ^2 Пирсона для распределения Коши будет обладать максимальной мощностью для близких альтернативных гипотез при разбиении на интервалы равной вероятности, когда потери информации, связанные с группированием исходной выборки, для распределения Коши будут минимальны.

ЛИТЕРАТУРА

1. Статистические методы в экспериментальной физике. М., Атомиздат, 1976.
2. Куллдорф Г. Введение в теорию оценивания по группированным и частично группированным выборкам. М., "Наука", 1966.
3. Бодин Н.А. Оценка параметров распределения по группированным выборкам. Тр. Матем. инст. АН СССР, III, 1970, 110-154.
4. Лемешко Б.Ю. Оценивание параметров распределений по группированным наблюдениям. - В сб.: Вопросы кибернетики, М., 1977, вып. 30, 80-96.
5. Губинский А.И., Денисов В.И., Гречко Ю.П., Лемешко Б.Ю., Цой Е.Б. Методические рекомендации по планированию экспериментов и обработке экспериментальных данных при исследовании надежности и качества функционирования систем "человек-техника". Л., ЛЭТИ, 1978.

С О Д Е Р Ж А Н И Е

Печатается в соответствии с решением Учёного совета факультета Автоматизированных систем управления Новосибирского электротехнического института от 13 декабря 1978 года.

печать 3.01.79г.

р. 3 Цена 1 руб 07 коп Зак. 38843

Производственно-издательский комбинат ВИНИТИ
Люберцы, Октябрьский пр., 403