УДК 519.232.2

В.Ю. Дубницкий, А.И. Ходырев

Харьковский институт банковского дела Университета банковского дела НБУ, Харьков

ОЦЕНИВАНИЕ ПАРАМЕТРОВ ДВУПАРАМЕТРИЧЕСКОГО РАСПРЕДЕЛЕНИЯ КОШИ МЕТОДОМ МАКСИМУМА ПРАВДОПОДОБИЯ ДЛЯ НЕГРУППИРОВАННЫХ И ГРУППИРОВАННЫХ НАБЛЮДЕНИЙ

Получены методом максимума правдоподобия оценки параметров положения и масштаба для распределения Коши. Приведены результаты численного эксперимента по определению этих параметров методом Ньютона. Качество полученных оценок оценено методом минимума хи-квадрат.

Ключевые слова: *распределение Коши, метод максимума правдоподобия, метод минимума хи-квадрат, численные методы, метод Ньютона, группированные наблюдения.*

Введение

Распределение Коши занимает особое место среди многочисленных функций распределения непрерывных случайных величин, известных в теории вероятностей.

Функция распределения Коши, согласно работе [1], имеет вид:

$$F(x)=\frac{1}{2}+\frac{1}{\pi}\operatorname{arctg}\frac{x-\mu}{\lambda},\ -\infty<x<\infty. \qquad (1)$$

Плотность распределения Коши имеет вид:

$$f(x)=\frac{\lambda}{\pi[\lambda^2+(x-\mu)^2]}. \qquad (2)$$

Входящие в условия (1) и (2) параметры λ и μ называют параметрами масштаба и положения λ соответственно. Параметр положения μ совпадает с модой и медианой распределения, параметр масштаба λ совпадает со срединным отклонением. Распределение вида (1)–(2) называют двухпараметрическим, приняв $\mu=1$ получаем однопараметрическое распределение.

В работе [2] для одномерной случайной величины X с симметричной плотностью распределения срединное отклонение E определено следующим образом:

$$P\{|X-\mu|<E\}=P\{|X-\mu|>E\}=\frac{1}{2}. \qquad (3)$$

Особенность распределения Коши в том, что оно не имеет ни начальных, ни центральных моментов вследствие расходимости определяющих их интегралов. В работе [3] отмечено, что это распределение относится к классу распределений с тяжелыми хвостами, то есть таких распределений, для которых не выполнен закон больших чисел. Это означает, что существуют выборочные оценки для среднего значения – величины $\overline{x}$ и среднего квадратического отклонения – величины σ, однако они ни к чему не сходятся и не имеют функциональной связи с параметрами распределения. В последние годы распределение Коши нашло применение в задачах оценки последствий стихийных бедствий [4], аварий сложных технических систем [5], обработки сигналов [6].

Анализ литературы. В работе [1] параметры положения μ и масштаба λ для упорядоченных и негруппированных выборок предложено определять методом квантилей.

Оценку $\hat{\mu}$ параметра положения μ получают из выражения:

$$\hat{\mu}=\frac{\hat{x}_{0,25}+\hat{x}_{0,75}}{2}. \qquad (4)$$

Оценку $\hat{\lambda}$ параметра масштаба λ получают из выражения

$$\lambda=\frac{\hat{x}_{0,75}-\hat{x}_{0,25}}{2}, \qquad (5)$$

где $\hat{x}_{0.25},\hat{x}_{0.75}$ – выборочные оценки первого и третьего квантилей.

Для негруппированной, но упорядоченной выборки выбирают значения вариационного ряда наблюдений с номерами $[(n+2)/4]$ и $[(3n+2)/4]$ соответственно.

В случае применения метода максимума правдоподобия (ММП) в работе [7] получено выражение для оценки параметра положения при условии, что параметр масштаба единичный, то есть для стандартной плотности распределения Коши, имеющей вид:

$$f(x)=\frac{1}{\pi[1+(x-\mu)^2]}. \qquad (6)$$

Для общего случая ($\lambda\neq1$) в доступной авторам данной работы литературе решение задачи оценки параметров с использованием ММП не найдено.

В некоторых экономических и физических задачах схема регистрации наблюдений такова, что фиксируют не конкретные численные значения признака, а только границы интервала, в которые попадает данный результат. Организованную таким образом выборку называют группированной. В работе [8] группированная выборка определена так.

Пусть область изменения признака известна из результатов наблюдения и совпадает с отрезком $[x_0, x_m]$. Разделим эту область на m интервалов, так, чтобы их границы отвечали условию $x_0 < x_1 < ... < x_{m-1} \le x_m$. В этом случае количество интервалов $J = m - 1$, а количество наблюдений в каждом из интервалов примем равным n_j. Учитывая способ формирования данных, применение ММП в этом случае имеет особенности, рассмотренные в работах [8, 9]. В этих же работах получены оценки параметров экспоненциального распределения, распределения Рэлея, Максвелла, гамма-распределения.

Постановка задачи. Пусть для выборки объёма n справедлив двухпараметрический закон распределения Коши. Требуется, используя ММП, получить оценки параметров положения и масштаба для несгруппированных и сгруппированных наблюдений.

Изложение результатов

Для негруппированных наблюдений, плотность распределения которых имеет вид (2), функция правдоподобия примет вид:

$$L = \frac{\lambda^n}{\pi^n} \prod_{i=1}^{n} \frac{1}{\lambda^2 + (x_i - \mu)^2}. \quad (7)$$

Логарифм функции правдоподобия примет вид:

$$\ln L = n \ln \lambda - n \ln \pi + \sum_{i=1}^{n} \ln \frac{1}{\lambda^2 + (x_i - \mu)^2}. \quad (8)$$

В соответствии с теорией ММП получим, что:

$$\frac{\partial \ln L}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} \frac{2\lambda}{\lambda^2 + (x_i - \mu)^2} = 0 \quad (9)$$

и

$$\frac{\partial \ln L}{\partial \mu} = 2\sum_{i=1}^{n} \frac{x_i - \mu}{\lambda^2 + (x_i - \mu)^2}. \quad (10)$$

Таким образом, оценки $\hat{\lambda}$ и $\hat{\lambda}$ параметров λ и μ будут корнями системы уравнений вида:

$$\begin{cases} \dfrac{n}{\lambda} - \displaystyle\sum_{i=1}^{n} \frac{2\lambda}{\lambda^2 + (x_i - \mu)^2} = 0; & (11.1) \\ \displaystyle\sum_{i=1}^{n} \frac{x_i - \mu}{\lambda^2 + (x_i - \mu)^2} = 0, & (11.2) \end{cases}$$

при условии, что $\lambda \ne 0$. Это условие в нашем случае выполняется всегда исходя из физического смысла задачи.

Для решения системы уравнений (11) в работе использован метод Ньютона в виде, изложенном в работе [10].

Обозначим условие (11.1) как $F(\lambda, \mu)$, условие (11.2) как $G(\lambda, \mu)$. Тогда система (11) примет вид:

$$\begin{cases} F(\lambda, \mu) = 0; \\ G(\lambda, \mu) = 0. \end{cases} \quad (12)$$

В соответствии с принятым вариантом метода на каждом $k -$ м шаге итерации выполняли следующие вычисления.

1. Вычисляли величины:

$$\begin{cases} \lambda_{k+1} = \lambda_k - \dfrac{\Delta_\lambda^{(k)}}{J(\lambda_k, \mu_k)}; \\ \mu_{k+1} = \mu_n - \dfrac{\Delta_\lambda^{(k)}}{J(\lambda_k, \mu_k)}. \end{cases} \quad (13)$$

2. Вычисляли якобиан системы (12):

$$J(\lambda_k, \mu_k) = \begin{vmatrix} F'_\lambda(\lambda, \mu) & F'_y(\lambda, \mu) \\ G'_\lambda(\lambda, \mu) & G'_\lambda(\lambda, \mu) \end{vmatrix}. \quad (14)$$

3. Вычисляли величину

$$\Delta_\lambda^{(n)} = \begin{vmatrix} F(\lambda_n, \mu_n) & F'_\mu(\lambda_n, \mu_n) \\ G(\lambda_n, \mu_n) & G'_\mu(\lambda_n, \mu_n) \end{vmatrix}. \quad (15)$$

4. Вычисляли величину

$$\Delta_\mu^{(n)} = \begin{vmatrix} F'_\lambda(\lambda_n, \mu_n) & F(\lambda_n, \mu_n) \\ G'_\lambda(\lambda_n, \mu_n) & G(\lambda_n, \mu_n) \end{vmatrix}. \quad (16)$$

5. Входящие в условия (13) – (16) частные производные функций $F(\lambda, \mu)$ и $G(\lambda, \mu)$ вычисляли, используя следующие выражения:

$$F_\lambda(\mu, \lambda) = 4\sum_{i=1}^{n} \frac{\lambda^2}{\left[\lambda^2 + \left(x_i - \mu^2\right)\right]^2} - \left(\frac{n}{\lambda^2} + 2\sum_{i=1}^{n} \frac{1}{\lambda^2 + (x_i - \mu)^2}\right); \quad (17)$$

$$F_\mu = 4\sum_{i=1}^{n} \frac{\lambda(\mu - x_i)}{\left[\lambda^2 + \left(x_i - \mu^2\right)\right]^2}; \quad (18)$$

$$G_\lambda(\mu, \lambda) = -2\lambda \sum_{i=1}^{n} \frac{x_i - \mu}{\left[\lambda^2 + (x_i - \mu)^2\right]^2}; \quad (19)$$

$$G_\mu(\mu, \lambda) = -2\sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\left[\lambda^2 + (x_i - \mu)^2\right]^2} - \sum_{i=1}^{n} \frac{1}{\lambda^2 + (x_i - \mu)^2}. \quad (20)$$

Для упрощения записи в выражениях (17) – (20) номер шага итерации равен k по умолчанию.

Пусть вероятность попадания наблюдения над случайной величиной с функцией распределения $F(x; \mu, \lambda)$ в интервал с номером j равна:

$$p_j(\mu, \lambda) = F(x_j; \mu, \lambda) - F(x_{j-1}; \mu, \lambda). \quad (21)$$

В условии (20) сохранены особенности символики, принятой в работах [8, 9]. Пусть количество наблюдений, попавших в j-й интервал, равно

$p_j^{nj}(\mu,\lambda)$. В этом случае, согласно работе [8], функция правдоподобия примет вид:

$$L_1 = \prod_{j=1}^{J} p_j^{nj}(\mu,\lambda). \quad (22)$$

Логарифм функции правдоподобия примет вид:

$$LnL_1 = \sum_{j=1}^{J}(nj)Ln(p_j(\mu,\lambda) \quad (23)$$

или:

$$LnL_1 = \sum_{j=1}^{J} nj(Ln((F(x_j;\mu,\lambda) - F(x_{j-1};\mu,\lambda)). \quad (24)$$

Условие (24) с учётом (21) примет вид:

$$LnL_1 = \sum_{j=1}^{J} nj(Ln(\tfrac{1}{\pi}(\mathrm{arctg}\tfrac{x_j-\mu}{\lambda} - \mathrm{arctg}\tfrac{x_{j-1}-\mu}{\lambda})). \quad (25)$$

В соответствии с теорией ММП для определения оценок $\hat{\mu}$ и $\hat{\lambda}$ необходимо решить следующую систему уравнений:

$$\begin{cases} \sum_{j=1}^{J} \dfrac{\left(nj\left(\mu^2 - \mu\left(x_j + x_{j-1}\right) + x_j x_{j-1} - \lambda^2\right)\left(x_j x_{j-1}\right)\right)}{A} = 0; \\ \sum_{j=1}^{J} \dfrac{nj\lambda\left(x_j - x_{j-1}\right)\left(x_j + x_{j-1} - 2\mu\right)}{A} = 0. \end{cases} \quad (26)$$

В системе (26) принято, что

$$A = \left(\left(x_j - \mu\right)^2 + \lambda^2\right)\cdot\left(\left(x_{j-1} - \mu\right)^2 + \lambda^2\right)\cdot \cdot\left(\mathrm{arctg}\frac{x_j-\mu}{\lambda} - \mathrm{arctg}\frac{x_j-\mu}{\lambda}\right). \quad (27)$$

Для решения этой системы используем приём, описанный в работе [11, С. 207-208]. Он позволяет свести задачу поиска решения системы нелинейных уравнений к задаче поиска минимума функции многих переменных. Обозначим уравнение (26.1) как $\varphi_1(\mu,\lambda)$ и уравнение (26.2) как $\varphi_2(\mu,\lambda)$.

Тогда решение системы (26) будет совпадать с решением задачи поиска безусловного минимума для функции вида:

$$\Phi(\mu,\lambda) = (\varphi_1(\mu,\lambda))^2 + (\varphi_2(\mu,\lambda))^2 \to \min_{\mu,\lambda}. \quad (28)$$

Поиск минимума в задаче (28) целесообразно выполнять, применяя поисковый алгоритм, предложенный в работе [12].

Для проверки предложенной в данной работе методики был проведен численный эксперимент.

На первом этапе эксперимента генерировали 250 квазислучайных чисел, распределённых по закону Коши с параметрами μ=2 и λ=1. Для получения массива М1 использовали соотношение, приведенное в работе [1]:

$$x_i = \mu + \lambda \mathrm{tg}(2\pi r_i), \quad (29)$$

где r_i=1,2,…250 – равномерно распределённое на отрезке [0;1] число. Для получения массива М2 использовали соотношение

$$x_i = \mu + \lambda\left(\frac{z_1}{z_2}\right), \quad (30)$$

где $Z_1 = Z_2 = N(0;1)$, то есть квазислучайные числа стандартного нормального распределения [1].

На втором этапе эксперимента полученную выборку упорядочивали по возрастанию и определяли оценки параметров μ и λ согласно выражению (5).

На третьем этапе эксперимента, используя полученные на втором этапе значения в качестве начальных, определяли оценки параметров закона распределения Коши, используя ММП в варианте, описанном в данной работе.

Второй и третий этапы осуществляли для каждой из выборок, полученных по условию (29) и (30).

На четвёртом этапе начальные значения параметров μ и λ получали для группированной выборки. Для этого использовали порядок вычислений, описанный в работе [13] и аналогичных учебных пособиях, после чего переходили ко второму этапу эксперимента. Для сравнения полученных результатов использовали критерий минимума χ^2, описанный в работе [14, С.30]. Результаты численного эксперимента приведены в таблице.

При анализе полученных результатов следует принять во внимание следующее. Количество интервалов, использованных при определении величины χ^2 и определённое по методике, изложенной в работе [13], равно девяти, число степеней свободы равно шести, α = 0,95.

Таблица

Результаты численного эксперимента по сравнению различных способов определения параметров распределения Коши

Способ формирования исходных данных	Вид выборки											
	Группированная						Негруппированная					
	Способ получения оценок						Способ получения оценок					
	Метод моментов			Метод максимума правдоподобия			Метод моментов			Метод максимума правдоподобия		
	μ	λ	χ^2	μ	λ	χ^2	μ	λ	χ^2	μ	λ	χ^2
М1	1,847	0,986	11,067	1,987	1,002	10,986	1,790	0,902	12,248	2,011	0,979	10,909
М2	1,851	0,938	11,073	1,963	0,965	10,865	1,802	0,915	12,306	1,956	1,092	10,808

Сравнивая полученные оценки, приходим к выводу, что оценки параметров распределения, полученные методом максимума правдоподобия, позволяют получить результаты, более точно определяющие параметры моделируемого распределения, чем метод моментов.

Выводы

1. Получены методом максимума правдоподобия оценки параметров распределения Коши для группированных и негруппированных выборок.

2. Предложен способ численного решения системы уравнений метода максимума правдоподобия для негруппированных выборок, реализующий метод Ньютона.

3. Предложен способ численного решения системы уравнений метода максимума правдоподобия для группированных выборок, использующий алгоритм поиска безусловного минимума.

4. Приведены результаты численного эксперимента, показавшего преимущества предложенного подхода.

Список литературы

1. Вадзинский Р.Н. Справочник по вероятностным распределениям ./ Р.Н. Вадзинский. – СПб.: Наука, 2001. – 295 с.

2. Справочник по вероятностным распределениям / Г.Г. Абезгауз, А.П. Тронь, Ю.Н. Копенки, И.А. Коровин. – М: Воениздат, 1966. – 408 с.

3. Родкин М.В. Теория экстремальных значений, распределения вероятностей для хвоста распределения для основных типов распределения, вопросы применения / М.В. Родкин. – М.: Международный институт теории прогноза землетрясений и математической геофизики РАН, 2002. – 27 с.

4. Кузнецов И.В. Методы расчёта ущерба от катастроф различного типа / И.В. Кузнецов, В.Ф. Писаренко, М.В. Родкин // Экономика и математические методы. – 1997. – Т. 33, вып. 4. – С. 39-50.

5. Аронов И.З. Методология оперативного управления безопасностью технических систем на основе процедур статистического анализа нарушений при эксплуатации и стандартизация методов оценки: автореф. дис. д-ра техн. наук / Аронов Иосиф Зиновьевич. – М., 1998.

6. Абрамов С.К. 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» / Абрамов Сергій Кирилович – Національний аерокосмічний університет ім. М.Є. Жуковського «ХАІ». – Х., 2003.

7. Кендалл М. Статистические выводы и связи / М. Кендалл, А. Стьюарт. – М.: Наука, 1973. – 900 с.

8. Куллдорф Г. Введение в теорию оценивания по группированным и частично группированным выборкам / Г. Кулдорф. – М.: Наука, 1966. – 176 с.

9. Лемешко Б.Ю. Об оценивании параметров распределения по группированным наблюдениям / Б.Ю. Лемешко // Вопросы кибернетики: Планирование эксперимента в прикладных исследованиях. Вып.30. – М.: Изд. АН СССР. Научный совет по комплексной проблеме «Кибернетика». – С. 80-96.

10. Копченова Н.В. Вычислительная математика в примерах и задачах / Н.В. Копченова, И.А. Марон. – М.: Наука, 1967. – 368 с.

11. Амосов А.А. Вычислительные методы для инженеров / А.А. Амосов, Ю.А. Дубинский, Н.В. Копченова. – М.: Высшая школа, 1994. – 544 с.

12. Дубницький В.Ю. Визначення параметрів виробничої функції із сталою еластичністю заміни / В.Ю. Дубницький, Б.В. Самородов // Системи обробки інформації. – Х.: Харківський університет Повітряних Сил, 2009. – Вип. 2 (76). – С. 109-112.

13. Практикум по теории статистики: Учеб. пособие / Под ред. Р.А. Шмойловой. – М.: Финансы и статистика, 2003. – 416 с.

14. Миленький А.В. Классификация сигналов в условиях неопределённости (Статистические методы самообучения в распознавании образов) / А.В. Миленький. – М.: Сов. радио, 1975. – 328 с.

Поступила в редколлегию 16.02.2011

Рецензент: д-р техн. наук, проф. С.Н. Герасин, Харьковский институт банковского дела Университета банковского дела НБУ, Харьков.

ОЦІНЮВАННЯ ПАРАМЕТРІВ ДВОПАРАМЕТРИЧНОГО РОЗПОДІЛУ КОШІ МЕТОДОМ МАКСИМУМУ ПРАВДОПОДІБНОСТІ ДЛЯ НЕГРУПОВАНИХ І ГРУПОВАНИХ СПОСТЕРЕЖЕНЬ

В.Ю. Дубницький, О.І. Ходирєв

Отримані методом максимуму правдоподібності оцінки параметрів положення і масштабу для розподілу Коші. Наведено результати чисельного експерименту для визначення цих параметрів методом Ньютона. Якість отриманих оцінок оцінена методом мінімуму хі-квадрат.

Ключові слова: *розподіл Коші, метод максимуму правдоподібності, метод мінімуму хі-квадрат, чисельні методи, метод Ньютона, груповані спостереження.*

EVALUATION OF CAUCHY TWO-PARAMETER DISTRIBUTION BY METHOD OF MAXIMUM LIKELIHOOD FOR GROUPED AND UNGROUPED OBSERVATIONS

V.Yu. Dubnitsky, A.I. Khodyrev

Evaluations of location and scope parameters for Cauchy distributions were obtained by method of maximum likelihood. The results are specified for a numerical experiment in determination of those parameters by Newton method. The quality of obtained evaluations was evaluated by minimum chi-square method.

Keywords: *Cauchy distribution, method of maximum likelihood, minimum chi-square method, numerical methods, Newton method, grouped observations.*