Записки научных
семинаров ПОМИ
Том 294, 2002 г.

В. В. Литвинова

ДВА КРИТЕРИЯ СОГЛАСИЯ ДЛЯ РАСПРЕДЕЛЕНИЯ КОШИ, ОСНОВАННЫЕ НА ХАРАКТЕРИЗАЦИЯХ

§1. Введение

Идея построения статистических критериев, основанных на характеризациях, восходит, по-видимому, к работе Ю. В. Линника [1]. Однако существует очень немного работ, в которых разрабатывались бы подобные критерии в применении к конкретным распределениям. В настоящей работе мы построим два критерии согласия для распределения Коши, основываясь на следующих характеризациях этого распределения.

Первая из них принадлежит Б. Арнольду [2] и формулируется так:

Пусть X и Y – независимые одинаково распределенные случайные величины (н.о.р.с.в.), имеющие непрерывную функцию распределения (ф.р.) F. Тогда X и $\frac{X+Y}{1-XY}$ одинаково распределены тогда и только тогда, когда F - ф.р. Коши с плотностью $f(x) = (\pi(1+x^2))^{-1}$.

Вторая характеризация восходит к работе Рамачандрана и Рао [3], см. также [2], теорема 13.7.2. Ее формулировка такова:

Пусть X и Y – н.о.р.с.в., имеющие непрерывную ф.р. F, а постоянные a и b таковы, что $|a|+|b| = 1$, причем числа $-\log|a|$ и $-\log|b|$ несоизмеримы.Тогда X и $aX + bY$ одинаково распределены тогда и только тогда, когда F – ф.р. Коши с произвольным параметром масштаба.

Пусть $X_1, ..., X_n, ...$ – независимые наблюдения, имеющие непрерывную ф.р. $G(x)$. Основываясь на первой характеризации, мы можем проверить простую основную гипотезу H_0 которая состоит в том, что G – функция распределения Коши F с плотностью f, против альтернативы H_1, состоящей в том, что $G(x) \neq$

Работа выполнена при поддержке гранта РФФИ 01-01-00245 и гранта научной школы 00-15-96019.

$F(x)$. Вторая характеризация позволяет проверить сложную гипотезу $H_0^{'}$ о том, что исходное распределение принадлежит семейству плотностей Коши с произвольным масштабом против альтернативы $H_1^{'}$, состоящей в том, что генеральная плотность не принадлежит этому семейству.

В соответствии с предложенными характеризациями, построим обычную эмпирическую функцию распределения:

$$G_n(t) = n^{-1} \sum_{i=1}^{n} \mathbf{I}\{X_i \leqslant t\}$$

и построим V-статистические функции распределения, соответствующие обеим характеризациям:

$$H_n(t) = n^{-2} \sum_{i,j=1}^{n} \mathbf{I}\Big\{ \frac{X_i + X_j}{1 - X_i X_j} \leq t \Big\},$$

$$L_n(t) = n^{-2} \sum_{i,j=1}^{n} \Big(\frac{1}{2}\mathbf{I}\Big\{ aX_i + bX_j \leqslant t \Big\} + \frac{1}{2}\mathbf{I}\Big\{ aX_j + bX_i \leqslant t \Big\}\Big).$$

Таким образом, критерии для проверки распределения Коши могут быть основаны на статистиках:

$$CH_n = \int\limits_{R^1} (H_n(t) - G_n(t))\, dG_n(t)$$

и

$$CL_n = \int\limits_{R^1} (L_n(t) - G_n(t))\, dG_n(t).$$

§2. Статистика CH_n

После некоторых преобразований легко увидеть, что статистика CH_n является V-статистикой степени 3 и ядром $\Phi_1(X, Y, Z)$:

$$\Phi_1(X, Y, Z) = \frac{1}{3}\mathbf{I}\Big\{ \frac{X+Y}{1-XY} \leqslant Z \Big\} + \frac{1}{3}\mathbf{I}\Big\{ \frac{Y+Z}{1-YZ} \leqslant X \Big\} + $$
$$+ \frac{1}{3}\mathbf{I}\Big\{ \frac{X+Z}{1-XZ} \leqslant Y \Big\} - \frac{1}{2}.$$

Предельные свойства этой статистики совпадают со свойствами U-статистики с тем же ядром:

$$KH_n = \frac{1}{C_n^3} \sum_{1 \leqslant i<j<k \leqslant n} \Phi_1(X_i, X_j, X_k)$$

Оказывается, что ядро $\Phi_1(X, Y, Z)$ является невырожденным. Действительно, найдем проекцию ядра

$$\psi_1(t) = \mathbf{E}(\Phi_1(X, Y, Z)|Z = t) \tag{1}$$

и вычислим дисперсию Δ_1^2 проекции при H_0:

$$\Delta_1^2 = \mathbf{E}\psi_1^2(Z). \tag{2}$$

Ясно, что

$$\psi_1(t) = \frac{1}{3}\mathbf{P}\Big\{\frac{X+Y}{1-XY} \leqslant t\Big\} + \frac{2}{3}\mathbf{P}\Big\{\frac{X+t}{1-Xt} \leqslant Y\Big\} - \frac{1}{2} = \frac{1}{3}F(t)+$$

$$+\frac{2}{3}\int\limits_{-\infty}^{\infty} dF(x) \int\limits_{\frac{x+t}{1-xt}}^{\infty} dF(y) - 1/2 = \frac{1}{3\pi}(\arctan t + 2\arctan\frac{1}{t}).$$

Следовательно,

$$\Delta_1^2 = \int\limits_{-\infty}^{\infty} \frac{1}{9\pi^2}(\arctan t + 2\arctan\frac{1}{t})^2 d(\frac{1}{\pi}\arctan t + \frac{1}{2}) = 1/36 > 0,$$

что и доказывает невырожденность ядра.

Применяя теорему Гефдинга [5, 6], получаем следующий результат.

Теорема 1. *Для статистики KH_n при $n \to \infty$ и справедливости H_0 имеет место соотношение:*

$$\sqrt{n}KH_n \xrightarrow{d} N(0, \frac{1}{4}).$$

Вычисление асимптотической эффективности статистики KH_n осложняется довольно громоздким видом видом ядра. Поэтому мы ограничимся лишь простейшим, но типичным случаем альтернативы сдвига, когда альтернативная функция распределения имеет вид $F(x-\theta)$, и вычислим бахадуровскую эффективность нашей статистики.

Обозначим через $c(\theta)$ бахадуровский наклон. Согласно [9, 10]:

$$c(\theta) = 2h(b(\theta)), \tag{3}$$

где

$$h(t) = t^2/(18\Delta^2)(1+o(1)), \qquad t \to 0 \tag{4}$$

$$b(\theta) = \mathbf{E}_\theta \Phi(X, Y, Z). \tag{5}$$

Для нашей статистики найдем $b(\theta)$.

$$b(\theta) = \mathbf{P}_\theta\Big\{\frac{X+Y}{1-XY} < Z\Big\} - \frac{1}{2} = \int\limits_{-\infty}^{\infty} dF(x-\theta) \int\limits_{-\infty}^{\infty} dF(y-\theta) \int\limits_{\frac{x+y}{1-xy}}^{\infty} dF(z-\theta)$$

$$-\frac{1}{2} = \int\limits_{-\infty}^{\infty} \frac{1}{\pi(1+(x-\theta)^2)} dx \int\limits_{-\infty}^{\infty} \frac{(1/2 - \frac{1}{\pi}\arctan(\frac{x+y}{1-xy} - \theta))}{\pi(1+(y-\theta)^2)} dy - \frac{1}{2}.$$

Вычисляя асимптотику этого интеграла при $\theta \to 0$, получаем:

$$b(\theta) \sim \frac{\theta}{\pi}.$$

Таким образом, локальный бахадуровский наклон статистики KH_n при альтернативе сдвига имеет вид:

$$c_{KH}(\theta) \sim \frac{4\theta^2}{\pi^2}, \qquad \theta \to 0.$$

Если определить в соответствии с [9] локальную бахадуровскую эффективность последовательности статистик $\{T_n\}$ как

$$\mathrm{eff}_T = \lim_{\theta \to o} c_T(\theta)/2K(\theta), \tag{6}$$

где $K(\theta) \sim \frac{1}{2} I_0 \theta^2, \theta \to 0$ и I_0 – информация Фишера в точке $\theta = 0$, то для альтернативы сдвига она составляет

$$\mathrm{eff}_{KH} = \frac{8}{\pi^2} \approx 0.8106,$$

что является довольно высоким значением.

§3. Статистика CL_n

Статистика CL_n, как и статистика CH_n, является V-статистикой степени 3 с ядром $\Phi_2(X,Y,Z)$:

$$\Phi_2(X,Y,Z)=\frac{1}{6}\mathbf{I}\Big\{aX+bY\leqslant Z\Big\}+\frac{1}{6}\mathbf{I}\Big\{aY+bX\leqslant Z\Big\}+\frac{1}{6}\mathbf{I}\Big\{aY+bZ\leqslant X\Big\}$$

$$+\frac{1}{6}\mathbf{I}\Big\{aZ+bY\leqslant X\Big\}+\frac{1}{6}\mathbf{I}\Big\{aX+bZ\leqslant Y\Big\}+\frac{1}{6}\mathbf{I}\Big\{aZ+bX\leqslant Y\Big\}-\frac{1}{2}$$

Предельные свойства этой статистики совпадают со свойствами U-статистики:

$$KL_n=\frac{1}{C_n^3}\sum_{1\leqslant i<j<k\leqslant n}\Phi_2(X_i,X_j,X_k).$$

Для невырожденного ядра $\Phi_2(X,Y,Z)$ найдем проекцию (1) и дисперсию проекции (2). Заметим, что если $\operatorname{sign}(a)=\operatorname{sign}(b)$, то наш критерий несостоятелен. Поэтому, не умаляя общности, будем считать, что $a>0$, $b<0$ (таким образом, $b=a-1$). Ясно, что

$$\psi_2(t)=\frac{1}{3}\mathbf{P}\Big\{aX+bY\leqslant t\Big\}+\frac{1}{3}\mathbf{P}\Big\{aX+bt\leqslant Y\Big\}+\frac{1}{3}\mathbf{P}\Big\{at+bY\leqslant X\Big\}-\frac{1}{2}.$$

Поскольку:
$\mathbf{P}\Big\{aX+bY\leqslant t\Big\}=F(t)$,
$\mathbf{P}\Big\{aX+bt\leqslant Y\Big\}=\mathbf{P}\Big\{Y-aX\geqslant bt\Big\}=1-F\Big(\frac{bt}{1+|a|}\Big)=1-F\Big(\frac{(a-1)t}{1+a}\Big)$,
$\mathbf{P}\Big\{at+bX\leqslant Y\Big\}=1-F\Big(\frac{at}{1+|b|}\Big)=1-F\Big(\frac{at}{2-a}\Big)$,
то для проекции получаем следующий результат:

$$\psi_2(t)=\frac{1}{3}F(t)-\frac{1}{3}F\big(\frac{(a-1)t}{1+a}\big)-\frac{1}{3}F\big(\frac{at}{2-a}\big)+\frac{1}{6}.$$

$$\Delta_2^2=\frac{1}{36}\int\limits_{-\infty}^{\infty}\Big(2F(t)-2F\Big(\frac{(a-1)t}{1+a}\Big)-2F\Big(\frac{at}{2-a}\Big)+1\Big)^2dF(t).$$

Для нахождения Δ_2^2 нам понадобится вычислить интегралы вида:
$h_1(\alpha):=\int_{-\infty}^{\infty}F(\alpha t)dF(t),\qquad \alpha\in R^1$,
$h_2(\alpha,\beta):=\int_{-\infty}^{\infty}F(\alpha t)F(\beta t)dF(t)$, $\alpha>0$, $\beta>0$.

Несложно получить, что для любого α

$$h_1(\alpha) = \frac{1}{2}.$$

Для вычисления $h_2(\alpha, \beta)$ воспользуемся леммой.

Лемма. $\int_{-\infty}^{\infty} F(\alpha t)F(\beta t)dF(t) = \frac{1}{3} - \frac{1}{2\pi^2}\Big(\mathrm{dilog}\Big(\frac{\alpha(\beta+1)}{\alpha+\beta}\Big)$
$+\mathrm{dilog}\Big(\frac{\beta(\alpha+1)}{\alpha+\beta}\Big) + \ln\Big(\frac{\alpha(\beta+1)}{\alpha+\beta}\Big)\ln\Big(\frac{\beta(\alpha+1)}{\alpha+\beta}\Big)\Big),$
где $F(t) = \frac{1}{2} + \frac{1}{\pi}\arctan t$.

Доказательство. Сначала продифференцируем этот интеграл по α. Получаем:

$$r(\alpha, \beta) := \Big(h_2(\alpha, \beta)\Big)'_\alpha = \int\limits_{-\infty}^{\infty} tf(\alpha t)f(t)F(\beta t)dt.$$

Теперь полученное выражение продифференцируем по β:

$$\Big(r(\alpha, \beta)\Big)'_\beta = \int\limits_{-\infty}^{\infty} t^2 f(\alpha t)f(t)f(\beta t)dt,$$

В предположении, что $\alpha > 0$ и $\beta > 0$, вычисляя $\Big(r(\alpha, \beta)\Big)'_\beta$, получаем:

$$\Big(r(\alpha, \beta)\Big)'_\beta = \frac{1}{\pi^2(\alpha+1)(\beta+1)(\alpha+\beta)}.$$

Интегрируя это выражение по β, имеем:

$$r(\alpha, \beta) + C_1 = \frac{\ln(\beta+1) - \ln(\alpha+\beta)}{\pi^2(\alpha^2-1)},$$

где константа C_1 находится из того, что $r(\alpha, 0) = 0$. Таким образом:

$$C_1 = -\frac{\ln\alpha}{\pi^2(\alpha^2-1)},$$

$$r(\alpha, \beta) = \frac{\ln(\beta+1) + \ln(\alpha) - \ln(\alpha+\beta)}{\pi^2(\alpha^2-1)}.$$

Найденное выражение для $r(\alpha,\beta)$ интегрируем по α и для краткости обозначим $D(x) := \mathrm{dilog} x = \int_1^x \frac{ln(t)}{1-t}dt$. Получаем:

$$h_2(\alpha,\beta)+C_2$$
$$=-\frac{1}{2\pi^2}\Big(D\Big(\frac{\alpha(\beta+1)}{\alpha+\beta}\Big)+D\Big(\frac{\beta(\alpha+1)}{\alpha+\beta}\Big)+\ln\Big(\frac{\alpha(\beta+1)}{\alpha+\beta}\Big)\ln\Big(\frac{\beta(\alpha+1)}{\alpha+\beta}\Big)\Big).$$

Поскольку $h_2(0,\beta)=0$, то

$$C_2=-\frac{1}{3},$$

$$h_2(\alpha,\beta)=\frac{1}{3}-\frac{1}{2\pi^2}\Big(D\Big(\frac{\alpha(\beta+1)}{\alpha+\beta}\Big)+D\Big(\frac{\beta(\alpha+1)}{\alpha+\beta}\Big)$$
$$+\ln\Big(\frac{\alpha(\beta+1)}{\alpha+\beta}\Big)\ln\Big(\frac{\beta(\alpha+1)}{\alpha+\beta}\Big)\Big).$$

Таким образом, применяя результаты вычислений $h_1(\alpha)$ и $h_2(\alpha,\beta)$, для дисперсии проекции получаем:

$$\Delta_2^2(a)=\frac{1}{108}-\frac{1}{18\pi^2}\Big(2\mathrm{dilog}\Big(\frac{1}{1+a}\Big)$$
$$+2D\Big(\frac{1}{2-a}\Big)+2D(1-a)-2D(a)-2D\Big(\frac{1-a}{a^2-a+1}\Big)-2D\Big(\frac{a}{a^2-a+1}\Big)$$
$$+\Big(\ln\frac{1}{1+a}\Big)^2+\Big(\ln\frac{1}{2-a}\Big)^2-\ln\Big(\frac{1-a}{a^2-a+1}\Big)\ln\Big(\frac{a}{a^2-a+1}\Big)\Big).$$

Как и ранее, используя теорему Гефдинга, получаем:

Теорема 2. *Для статистики KL_n при $n\to\infty$ справедливо:*

$$\sqrt{n}KG_n \xrightarrow{d} N(0,9\Delta_2^2(a)).$$

Вычислим локальный бахадуровский наклон и локальную эффективность нашей статистики в предположении, что X_i при альтернативе имеют функцию распределения $F(x-\theta)$ и используя (3) – (6). Имеем

$$b(\theta)=\mathbf{P}_\theta\Big\{aX+bY\leqslant Z\Big\}-\frac{1}{2}=$$

$$=\mathbf{P}_\theta\Big\{a(X-\theta)+b(Y-\theta)-(Z-\theta)\leqslant\theta-a\theta-b\theta\Big\}-\frac{1}{2}=$$

$$= F\Big(\frac{\theta(1-a-b)}{2}\Big) - \frac{1}{2} = F(\theta(1-a)) - \frac{1}{2}.$$

Таким образом, для бахадуровского наклона получили:

$$c_{KL}(\theta) \sim \frac{(1-a)^2\theta^2}{9\pi^2\Delta_2^2(a)}, \qquad a \in (0,1).$$

Следовательно, локальная бахадуровская эффективность составляет:

$$\mathrm{eff}^B_{KG}(a) = 8(1-a)^2 \Big/ \Big[\frac{1}{3} - \frac{2}{\pi^2}\Big(2D\Big(\frac{1}{1+a}\Big) + 2D\Big(\frac{1}{2-a}\Big) +$$

$$+2D(1-a) - 2D(a) - 2D\Big(\frac{1-a}{a^2-a+1}\Big) - 2D\Big(\frac{a}{a^2-a+1}\Big) +$$

$$+\Big(\ln\frac{1}{1+a}\Big)^2 + \Big(\ln\frac{1}{2-a}\Big)^2 - \ln\Big(\frac{1-a}{a^2-a+1}\Big)\ln\Big(\frac{a}{a^2-a+1}\Big)\Big)\Big].$$

Исследование $\mathrm{eff}^B_{KG}(a)$ как функции от a с помощью пакета *Maple* 6 показывает, что $\mathrm{eff}_{KG}(a)$ имеет выпуклый график на отрезке $[0,1]$. Максимум достигается при $a = 0.3730$ и составляет 0.6850. Для построения критерия следует выбрать a, близкое к указанному значению и такое, чтобы выполнялось условие несоизмеримости $-\log a$ и $-\log(1-a)$.

Литература

1. Ю. В. Линник, *Линейные формы и статистические критерии.* — Украинский матем. журнал **5**, № 2 (1953), 207–243 no. 3, 247–290.
2. B. Arnold, *Some characterizations of the Cauchy distribution..* — Austr. J. Stat. **21** (1979), 166–169.
3. B. Ramachandran, C. R. Rao, *Solutions of functional equations arising in some regression problems and a characterization of the Cauchy law.* — Sankhya **32, Ser. A, No. 1** (1970), 1–30.
4. А. М. Каган, Ю. В. Линник, С. Р. Рао, *Характеризационные задачи математической статистики.* М., Наука, 1972.
5. Королюк В. С., Боровских Ю. В., *Теория U-статистик.* Киев, Наукова Думка, 1989.
6. Hoeffding W., *A class of statistics with asymptotically normal distribution..* — Ann. Math. Statist. **19** (1948), 293–325.
7. Рао С. Р., *Линейные статистические методы и их приложения.* пер. с англ., М., Наука, 1973.
8. Rao C. R., *Criteria of estimation in large samples.* — Sankhya **A25** (1963), 189–206.
9. Bahadur R. R., *Some limit theorem in statistics.* — SIAM Philadelphia (1971).

10. Никитин Я. Ю.,Поникаров Е. В., *Грубая асимптотика вероятностей больших уклонений черновского типа функционалов Мизеса и U–статистик.* — Труды СПбМО **7** (1999).

Litvinova V. V. Two criteria of goodness of fit for Cauchy distributions based on characterizations.

In this work we construct two criteria of goodness of fit for Cauchy distribution, based on the characterizations. The first belongs to Arnold and the second mentioned in the Ramachandran's paper. The Bahadur efficiency for these criteria are fined.

С.-Петербургский
государственный университет

Поступило 10 июля 2002 г.