Записки научных
семинаров ПОМИ
Том 351, 2007 г.

В. Б. Невзоров

ЭКСТРЕМУМЫ И t_2–РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА

В статистическом мире хорошо известно семейство распределений Стьюдента. Говорят, что случайная величина X имеет t_ν-распределение (или распределение Стьюдента с ν степенями свободы, где $\nu > 0$), если ее плотность распределения задается соотношением

$$f_\nu(x) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}} \left(1 + \frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \ -\infty < x < \infty, \qquad (1)$$

где $\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}dx,\ s > 0$, обозначает гамма-функцию. Это семейство распределений в начале прошлого века было введено в статистическую теорию и практику Уильямом Госсетом, подписывавшим свои математические работы псевдонимом Student. В начале XXI века появилась серия статей ([1]–[5]), в которых авторы обратили внимание читателей на ряд интересных свойств одного из представителей этого семейства t_2-распределения с плотностью

$$f_2(x) = (2 + x^2)^{-3/2} \qquad (2)$$

и функцией распределения

$$F_2(x) = \frac{1}{2}\left(1 + \frac{x}{\sqrt{2+x^2}}\right), \quad -\infty < x < \infty. \qquad (3)$$

В частности, были получены некоторые характеризации t_2-распределения свойствами порядковых статистик. В данной работе предлагаются две новые характеризации этого типа распределений.

Порядковые статистики и t_2-распределение

Пусть $X,\ X_1,\ X_2, \ldots$ – независимые одинаково распределенные случайные величины (с.в.) с общей функцией распределения

Работа частично поддержана грантами РФФИ 07-01-00688 и 06-01-00179.

$F(x)$, плотностью распределения $f(x)$ и конечным математическим ожиданием, а $X_{1,n} \le X_{2,n} \le \ldots \le X_{n,n}, n = 1, 2, \ldots,$ – соответствующие порядковые статистики. Дополнительно предположим, что $f(x) > 0$, если $\gamma < x < \delta$, где $\gamma = \inf\{x : F(x) > 0\}$ и $\delta = \sup\{x : F(x) < 1\}$. В работе [1] было показано, что эквивалентными являются следующие утверждения:

a) $E(X_{1,3} + X_{3,3}|X_{2,3} = x) = 2x$ п.н.; (4)

b) $E(X_1 + X_2 + X_3|X_{2,3} = x) = 3x$ п.н.; (5)

c) $(F(x)(1 - F(x))^{3/2} = cf(x), \quad \gamma < x < \delta, \quad c > 0;$ (6)

d) $\gamma = -\infty,\ \delta = \infty$ и $F(x) = F_2((x - \mu)/\sigma),\ -\infty < \mu < \infty,\ \sigma > 0,$ (7)

где функция распределения F_2 определена в (3).

Вслед за [1] появился ряд работ ([2]–[5]), в которых были получены новые характеризации t_2-распределения, а также различные обобщения приведенного выше результата на случай выборок объема $n = 2k + 1,\ k = 1, 2, \ldots$. Например, в работе [5] соотношения (4)-(7) были дополнены, как оказалось, эквивалентными им утверждениями вида:

e) $E(X_1X_2X_3|X_{2,3} = x) = cx$ п.н.; (8)

f) $E(X_{1,3}X_{3,3}|X_{2,3} = x) = c$ п.н., (9)

где c – некоторая отрицательная константа. В работе [2] было показано, что t_2-распределение можно, в частности, характеризовать соотношением

$$E(X_1 + \ldots + X_{2k+1}|X_{k+1,2k+1} = x) \\ = \frac{2k+1}{2m+1} E(X_1 + \ldots + X_{2m+1}|X_{m+1,2m+1} = x), \quad (10)$$

выполняемым для некоторых $k \ne m$.

Экстремумы и t_2-распределение

Все порядковые статистики, приведенные в равенствах (4) и (9), строились по одной и той же выборке $X_1,\ X_2,\ X_3$. Рассмотрим теперь несколько иную ситуацию. Вместо порядковых статистик $X_{1,3},\ X_{2,3}$ и $X_{3,3}$, участвовавших в регрессионных соотношениях (4) и (9), обратимся к максимумам

$$M_1 = X_1 = X_{1,1}, \ M_2 = \max\{X_1, X_2\} = X_{2,2}$$
$$\text{и} \quad M_3 = \max\{X_1, X_2, X_3\} = X_{3,3}.$$

Выясняется, что t_2-распределение можно характеризовать соотношением, аналогичным равенству (4), в котором только вместо порядковых статистик $X_{1,3}$, $X_{2,3}$, $X_{3,3}$ берутся максимальные порядковые статистики $X_{1,1}$, $X_{2,2}$ и $X_{3,3}$. Справедлив следующий результат.

Теорема 1. *Пусть X_1, X_2, X_3 – независимые одинаково распределенные с.в. с функцией распределения $F(x)$, плотностью распределения $f(x)$ и конечным математическим ожиданием. Пусть также $f(x) > 0$, если $\gamma < x < \delta$, где $\gamma = \inf\{x : F(x) > 0\}$ и $\delta = \sup\{x : F(x) < 1\}$. Тогда эквивалентными являются утверждения:*

$$E(M_1 + M_3 | M_2 = x) = \frac{3}{2}x + a \quad \text{п.н., где } a \text{ – некоторая константа}; \tag{11}$$

и

$$\gamma = -\infty, \ \delta = \infty \text{ и } F(x) = F_2((x - 2a)/\sigma), \ \sigma > 0, \tag{12}$$

где функция распределения F_2 определена в (3).

Замечание. Видим, что приведенная характеризация t_2-распределения имеет ряд особенностей, отличающих ее от характеризации этого же распределения регрессионными свойствами порядковых статистик $X_{1,3}$, $X_{2,3}$, $X_{3,3}$. Во-первых, отличаются коэффициенты перед x в правых частях соотношений (4) и (11). Во вторых, если соотношение (4) справедливо для всех представителей семейства распределений t_2, т.е. оно не реагирует на сдвиг и изменение масштаба, то соотношение (11) также не зависит от изменения масштаба, но зато реагирует на сдвиг.

Свойства порядковых статистик $X_{1,3}$, $X_{2,3}$, $X_{3,3}$ и максимумов M_1, M_2, M_3 по их отношению к t_2 – распределению могут быть объединены следующей характеризационной теоремой.

Теорема 2. *В условиях теоремы* 1 *при дополнительном предположении, что исходные с.в. имеют нулевое математическое ожидание, эквивалентны утверждения:*

$$E(M_1 + M_3 | M_2 = x) = \frac{3}{4}E(X_{1,3} + X_{3,3} | X_{2,3} = x) \quad \text{п.н.}; \tag{13}$$

и

$$\gamma = -\infty,\ \delta = \infty \quad \text{и} \quad F(x) = F_2(x/\sigma),\ \sigma > 0. \tag{14}$$

Доказательство теоремы 1. Нам потребуются следующие достаточно очевидные соотношения:

$$E(M_1|M_2 = x) = \frac{x}{2} + \int_{-\infty}^{x} udF(u)/2F(x); \tag{15}$$

$$E(M_3|M_2 = x) = xF(x) + \int_{x}^{\infty} udF(u); \tag{16}$$

$$E(X_{1,3}|X_{2,3} = x) = \int_{-\infty}^{x} udF(u)/F(x); \tag{17}$$

$$E(X_{3,3}|X_{2,3} = x) = \int_{x}^{\infty} udF(u)/(1 - F(x)). \tag{18}$$

Предположим, что $EX = \mu$ и обозначим

$$I(x) = \int_{-\infty}^{x} udF(u). \tag{19}$$

Тогда, учитывая, что

$$\mu = I(x) + \int_{x}^{\infty} udF(u),$$

равенство (4) можно записать в виде

$$\frac{I(x)}{F(x)} + \frac{\mu - I(x)}{1 - F(x)} = 2x,\ \gamma < x < \delta, \tag{20}$$

что равносильно соотношению

$$I(x) = \frac{2xF(x)(1 - F(x))}{1 - 2F(x)} - \frac{\mu F(x)}{1 - 2F(x)},\ \gamma < x < \delta. \tag{21}$$

В работе [1] было показано, что утверждение (4) равносильно утверждению (7). Следовательно, если функция распределения

F удовлетворяет соотношению (21) при некотором μ, то она обязана иметь вид

$$F(x) = F_2((x-\mu)/\sigma), \ \sigma > 0.$$

Рассмотрим теперь равенство (11). Оно сводится к соотношению

$$\frac{x}{2} + \int\limits_{-\infty}^{x} udF(u)/2F(x) + xF(x)$$

$$+ \int\limits_{x}^{\infty} udF(u) = \frac{3}{2}x + a, \ \gamma < x < \delta. \tag{22}$$

Вновь обозначая

$$I(x) = \int\limits_{-\infty}^{x} udF(u) \quad \text{и} \quad \mu = EX = \int\limits_{-\infty}^{\infty} udF(u),$$

получаем из (22) уравнение (относительно $F(x)$)

$$I(x) = \frac{2xF(x)(1-F(x))}{1-2F(x)} - \frac{2(\mu-a)F(x)}{1-2F(x)}, \quad \gamma < x < \delta. \tag{23}$$

Сравнивая (23) с (21), выводим, что решением этого уравнения является любая функция распределения вида

$$F(x) = F_2\left((x-2(\mu-a))/\sigma\right), \ \sigma > 0, \tag{24}$$

и $\gamma = -\infty$, $\delta = \infty$. Поскольку математическое ожидание с.в. X, имеющей функцию распределения (24), равно $2(\mu-a)$, а у нас еще есть предположение (точнее, обозначение), что $EX = \mu$, то видим, что μ должно удовлетворять соотношению $\mu = 2(\mu-a)$, т.е. $\mu = 2a$. Таким образом, при любом фиксированном значении a соотношение (11) справедливо только для функций распределения

$$F(x) = F_2\left((x-2a)/\sigma\right), \ \sigma > 0,$$

что и доказывает теорему 1.

В частности, видим, что равенство

$$E(M_1 + M_3 | M_2 = x) = \frac{3}{2}x$$

характеризует семейство функций распределения

$$F(x) = F_2(x/\sigma), \ \sigma > 0.$$

Доказательство теоремы 2. Используя соотношения (15)–(18), равенство (13) (учитывая, что $EX = 0$) можно переписать в виде

$$4\left(\frac{x}{2} + \frac{I(x)}{2F(x)} + xF(x) - I(x)\right) = 3\left(\frac{I(x)}{F(x)} - \frac{I(x)}{1-F(x)}\right), \quad \gamma < x < \delta. \tag{25}$$

Из (25) получаем уравнение относительно $F(x)$ вида

$$I(x) = \frac{2xF(x)(1-F(x))}{1-2F(x)}, \quad \gamma < x < \delta. \tag{26}$$

Сравнивая (26) с (21), получаем, что $\gamma = -\infty$, $\delta = \infty$ и решением уравнения (26) являются функции распределения

$$F(x) = F_2(x/\sigma), \quad \sigma > 0.$$

Литература

1. В. Б. Невзоров, *Об одном свойстве распределения Стьюдента с двумя степенями свободы.* — Зап. научн. семин. ПОМИ **294** (2002), 148–157.
2. В. Б. Невзоров, *О характеризациях распределения Стьюдента с двумя степенями свободы.* — Вестник СПбГУ, сер.1 (2005), вып. 4, 38–42.
3. V. B. Nevzorov, N. Balakrishnan, and M. Ahsanullah, *Simple characterization of Student's t_2 – distribution.* — The Statistician, **52** (2003), part 3, 395–400.
4. I. S. Akhundov, N. Balakrishnan, and V. B. Nevzorov, *New characterizations by properties of midrange and related statistics.* — Communications in statistics. Theory and method, **33** (2004), No. 12, 3133–3143.
5. L. Nevzorova, V. Nevzorov, and I. Akhundov, *A simple characterization of Student's t_2 – distribution.* — Metron-International journal of statistics, v. LXV (2007), No. 1, 53–57.

Nevzorov V. B. Extremes and Student's t_2 -distribution.

Some new characterizations of Student's distribution on two degrees of freedom by regressional properties of maximal order statistics are obtained.

С.-Петербургский государственный университет

E-mail: Valnev@mail.ru

Поступило 30 ноября 2007 г.