Записки научных
семинаров ПОМИ
Том 361, 2008 г.

Н. М. Марудова, В. Б. Невзоров

ОБ ОДНОМ СЕМЕЙСТВЕ, СОДЕРЖАЩЕМ t_2-РАСПРЕДЕЛЕНИЕ СТЬЮДЕНТА

В 2002 году почти одновременно появились две статьи ([1] и [2]), в которых главным объектом исследований стал один из самых простейших представителей семейства распределений Стьюдента.

Случайная величина (с.в.) X имеет t_ν-распределение Стьюдента (распределение Стьюдента с ν степенями свободы, $\nu = 1, 2, \dots$), если ее плотность распределения имеет вид

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

где $\Gamma(s) = \int\limits_0^\infty x^{s-1} \exp(-x)\, dx$, $s > 0$, обозначает гамма-функцию. В частности, t_2-распределение имеет плотность

$$f_2(x) = \left(2 + x^2\right)^{-3/2} \tag{2}$$

и функцию распределения

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

Джоунс [1] подошел к t_2-распределению не как к представителю семейства распределений, широко используемого в математической статистике, а с методологической точки зрения. Он привел достаточно убедительные аргументы, доказывающие, что t_2-распределение (наряду с распределениями Коши, Лапласа, логистическим и нормальным) следует использовать в элементарных вероятностных курсах для иллюстрации различных характеристик и свойств непрерывных распределений. В работе же [2] было показано, что t_2-распределение

Работа частично поддержана грантом РФФИ 07-01-00688 .

может не только служить хорошим учебным примером при изучении непрерывных распределений, но и играет существенную роль в теории порядковых статистик. Одним из первых результатов, характеризующих t_2-распределение, был следующий.

Пусть X_1, X_2, X_3 – независимые случайные величины, имеющие функцию распределения $F(x)$ и плотность распределения $f(x)$, а $X_{1,3} \le X_{2,3} \le X_{3,3}$ – соответствующие порядковые статистики. Пусть также $\bar{X} = (X_1 + X_2 + X_3)/3$ – выборочное среднее и $R = (X_{1,3} + X_{3,3})/2$ – выборочная середина размаха. Тогда эквивалентными являются следующие утверждения:

(a)
$$E\left(R | X_{2,3} = x\right) = x \quad \text{п.н.}; \tag{4}$$

(b)
$$E\left(\bar{X} | X_{2,3} = x\right) = x \quad \text{п.н.}; \tag{5}$$

(c)
$$E\left(X_1 | X_{2,3} = x\right) = x \quad \text{п.н.}; \tag{6}$$

(d)
$$\left(F(x)\left(1 - F(x)\right)\right)^{3/2} = cf(x), \tag{7}$$

где c – некоторая положительная константа;

(e)
$$F(x) = F_2\left((x - \mu)/\sigma\right), \quad -\infty < \mu < \infty, \sigma > 0,$$

где F_2 определена в (3).

Таким образом, были получены характеризации t_2-распределения регрессионными соотношениями, включающими выборочную медиану $X_{2,3}$, выборочное среднее $\bar{X}$, выборочную середину размаха R. Там же в [2] (для порядковых статистик

$$X_{1,2k+1} \le \cdots \le X_{k,2k+1} \le X_{k+1,2k+1} \le X_{k+2,2k+1} \le \cdots \le X_{2k+1,2k+1},$$

выборочной медианы $X_{k+1,2k+1}$ и выборочной середины квазиразмаха $R_k = (X_{k,2k+1} + X_{k+2,2k+1})/2$ уже при произвольном $k = 1, 2, \ldots$) было рассмотрено несколько более общее (чем (4)) соотношение

$$E\left(R_k | X_{k+1,2k+1} = x\right) = x \quad \text{п.н.} \tag{8}$$

и было показано, что (8) характеризует некоторое распределение с функцией распределения $F(x)$, удовлетворяющей при некотором $c > 0$ равенству

$$(F(x)(1 - F(x)))^{\frac{k}{2}+1} = cf(x). \tag{9}$$

Таким образом, выяснилось, что регрессионные соотношения типа равенства (8) характеризуют ряд распределений, принадлежащих семейству, функции распределения которого F определяются равенствами

$$(F(x)(1 - F(x)))^{\alpha} = cF'(x), \tag{10}$$

где c – некоторая положительная константа и $-\infty < \alpha < \infty$.

Интересно отметить, что семейство распределений (10) кроме уже упомянутого выше t_2-распределения (которое соответствует $\alpha = 3/2$) содержит и такие типы классических распределений, как логистическое (при $\alpha = 1$) и равномерное (при $\alpha = 0$).

Характеризация t_2-распределения свойством (4) привела к несколько более общему результату, использующему соотношения (8) при различных целых значениях k, что дает основания родственными для распределения Стьюдента с двумя степенями свободы считать все распределения, удовлетворяющие равенству (9) при $k = 2, 3, \ldots$.

Несколько позже в работе [3] была получена еще одна характеризация t_2-распределения, связанная с регрессионными свойствами упорядоченных случайных величин. На этот раз по независимым с.в. X_1, X_2, X_3 строились последовательные максимумы $M_1 = X_1, M_2 = \max\{X_1, X_2\}, M_3 = \max\{X_1, X_2, X_3\}$. Было показано, что эквивалентными являются условия:

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

где a – некоторая константа, и

$$F(x) = F_2((x - 2a)/\sigma), \quad \sigma > 0, \tag{12}$$

где функция распределения F_2 определена в (3).

Естественным является желание посмотреть, какие распределения характеризуются соотношениями вида (11), если в них рассматривать последовательные максимумы

$$M_k = \max\{X_1, X_2, \ldots, X_k\}$$

для выборок большего объема. Справедлив следующий результат.

Теорема 1. *Пусть $X_1, X_2, \ldots, X_{k-1}, X_k, X_{k+1}$ – независимые одинаково распределенные случайные величины с конечным математическим ожиданием, функцией распределения $F(x)$ и плотностью распределения $f(x)$. Пусть также $f(x) > 0$, если $\gamma < x < \delta$, где $\gamma = \inf\{x : F(x) > 0\}$ и $\delta = \sup\{x : F(x) < 1\}$. Тогда для любого фиксированного $k = 2, 3, \ldots$ из равенства*

$$E(M_{k-1} + M_{k+1} | M_k = x) = \left(2 - \frac{1}{k}\right)x + a, \quad \gamma < x < \delta, \quad \text{п.н.}, \tag{13}$$

где a – некоторая константа, следует, что $\gamma = -\infty$, $\delta = \infty$ и

$$G(x) = c \int_{1/2}^{x} (1-u)^{\left(\frac{1}{k}-2\right)} u^{\left(1-k-\frac{1}{k}\right)} du + d, \quad 0 < x < 1, \tag{14}$$

где $G(x) = F^{-1}(x)$ – функция обратная функции распределения F, $c > 0$ и d – некоторые константы.

Замечание 1. Нетрудно убедиться, что если (13) справедливо для с.в. $X_1, X_2, \ldots, X_{k+1}$ с функцией распределения $F(x)$, то при переходе к величинам $Y_r = \alpha X_r + \beta$, $r = 1, 2, \ldots, k+1$, $\alpha > 0$, $-\infty < \beta < \infty$, т.е. к случайным величинам с функцией распределения $F((x-\beta)/\alpha)$, принадлежащим тому же самому семейству распределений, что и исходные случайные величины, мы остаемся в рамках соотношения (13), заменив лишь константу a на другую константу $\alpha a + \beta/k$. Тогда в соотношении (14) константа c заменится на αc и вместо d нужно будет взять $\alpha d + \beta$. Сравнительно просто также убедиться, что если исходная функция распределения F задается своей обратной функцией G, имеющей вид (14), то тогда будет выполнено равенство (13) с некоторой константой a, зависящей от c и d. Все это означает, что семейство функций распределения $\{F((x-\beta)/\alpha), \alpha > 0, -\infty < \beta < \infty\}$, определяемых своими обратными функциями G, которые имеют вид (14), характеризуется равенством (13), в котором а может принимать любые значения ($-\infty < \alpha < \infty$).

Замечание 2. Простые выкладки показывают, что если функция $G(x)$ удовлетворяет соотношению (14), то справедливо равенство

$$(F(x))^{k-1+\frac{1}{k}} (1 - F(x))^{2-\frac{1}{k}} = cF'(x), \quad -\infty < x < \infty, \tag{15}$$

с той же константой c, что и в соотношении (14). Представляет интерес сравнение равенств (15) при $k = 2, 3, \dots$ с аналогичным набором соотношений (9).

Доказательство теоремы 1. В замечании 1 было отмечено, что за счет сдвига исходных X-ов (рассматривая, например, $Y_r = X_r - ka$, $r = 1, 2, \dots, k+1$) можно перейти к равенству (13), в котором $a = 0$. Поэтому для простоты изложения мы с самого начала рассмотрим соотношение

$$E\left(M_{k-1} + M_{k+1} | M_k = x\right) = \left(2 - \frac{1}{k}\right)x. \tag{16}$$

Нетрудно проверить, что совместное распределение максимумов M_k и M_r $(r > k)$ имеет вид

$$P\left\{M_k < x, M_r < y\right\} = F^r\left(y\right), \tag{17}$$

если $y < x$, и

$$P\left\{M_k < x, M_r < y\right\} = F^k\left(x\right) F^{r-k}\left(y\right), \tag{18}$$

если $y \geq x$.

Отсюда следует, что при $r > k$ справедливы равенства

$$E\left(M_r | M_k = x\right) = xF^{r-k}\left(x\right) + (r-k) \int\limits_{F(x)}^{1} G\left(y\right) y^{r-k-1} dy \tag{19}$$

и

$$E\left(M_k | M_r = x\right) = \frac{k}{r}x + \frac{(r-k)\,k}{rF^k\left(x\right)} \int\limits_{0}^{F(x)} G\left(y\right) y^{k-1} dy. \tag{20}$$

В частности, получаем, что

$$E\left(M_{k-1} | M_k = x\right) = \frac{k-1}{k}x + \frac{(k-1)}{kF^{k-1}\left(x\right)} \int\limits_{0}^{F(x)} G\left(y\right) y^{k-2} dy \tag{21}$$

и

$$E\left(M_{k+1} | M_k = x\right) = xF\left(x\right) + \int\limits_{F(x)}^{1} G\left(y\right) dy. \tag{22}$$

Используя равенства (21) и (22), можно соотношение (16) переписать в виде

$$\frac{k-1}{k}x + \frac{(k-1)}{kF^{k-1}(x)} \int\limits_0^{F(x)} G(y)\, y^{k-2} dy + xF(x) + \int\limits_{F(x)}^1 G(y)\, dy = \Big(2 - \frac{1}{k}\Big)x. \quad (23)$$

Из (23) следует, что

$$(k-1)\int\limits_0^z G(y)\, y^{k-2} dy + kz^k G(z) + kz^{k-1} \int\limits_z^1 G(y)\, dy = kz^{k-1} G(z), \quad 0 < z < 1. \quad (24)$$

После дифференцирования левой и правой частей (24) по z и простых преобразований полученного равенства получаем, что

$$k(k-1)\int\limits_z^1 G(y)\, dy = G(z)\Big(kz - k^2 z + (k-1)^2\Big) + G'(z)\left(kz - kz^2\right). \quad (25)$$

Повторное дифференцирование приводит к уравнению

$$\left(k^2 z + kz - k^2 + k - 1\right) G'(z) = \left(kz - kz^2\right) G''(z), \quad (26)$$

решение которого можно записать в виде

$$G(z) = c\int\limits_{1/2}^z (1-u)^{\frac{1}{k}-2} u^{\left(1-k-\frac{1}{k}\right)} du + d, \quad (27)$$

где $c > 0$ и $-\infty < d < \infty$ – произвольные константы. Из (27) вытекает, что $\gamma = G(0) = -\infty$ и $\delta = G(1) = \infty$. Таким образом, теорема доказана.

В теореме 1 мы рассмотрели простейшее обобщение соотношения (11) и получили некоторое новое расширение множества распределений (задаваемое равенствами (14) при $k = 2, 3, \ldots$), содержащих t_2-распределение. Возникает вопрос, можно ли существенно расширить семейство распределений (14), если рассмотреть регрессионные соотношения, аналогичные (13), в более общей постановке.

Рассмотрим максимальные порядковые статистики

$$M_{k-j} = \max\{X_1, X_2, \ldots, X_{k-j}\}, \quad M_k = \max\{X_1, X_2, \ldots, X_k\}$$

и M_{k+r} для произвольных k, j и r таких, что $k - j \geq 1$ и $r \geq 1$. Справедлив следующий результат.

Теорема 2. *Пусть $X_1, X_2, \ldots, X_k, \ldots, X_{k+r}$ – независимые одинаково распределенные случайные величины с конечным математическим ожиданием, функцией распределения $F(x)$ и плотностью распределения $f(x)$. Пусть также $f(x) > 0$, если $\gamma < x < \delta$, где $\gamma = \inf\{x : F(x) > 0\}$ и $\delta = \sup\{x : F(x) < 1\}$. Тогда для любых фиксированных целых k, j и r таких, что $k \geq j + 1$ и $r \geq 1$ из равенства*

$$E(M_{k-j} + M_{k+r} | M_k = x) = \Bigl(2 - \frac{j}{k}\Bigr)x + a, \quad \gamma < x < \delta \ \text{ п.н.}, \tag{28}$$

где a – некоторая константа, следует, что $\gamma = -\infty, \delta = \infty$ и

$$G(x) = c\int\limits_{1/2}^{x} (1 - z^r)^{\mu} z^{\nu} dz + d, \quad 0 < x < 1, \tag{29}$$

где

$$\mu = \frac{j^2}{kr} - 1 - \frac{j}{r}, \quad \nu = 2j - \frac{j^2}{k} - k - 1,$$

и $G(x) = F^{-1}(x)$ – функция обратная функции распределения F, а $c > 0$ и d – некоторые константы.

Замечание 3. Как частный случай из теоремы 2 при $j = r = 1$ следует теорема 1.

Замечание 4. Функция распределения F, задаваемая в (29) своей обратной функцией G, удовлетворяет соотношению

$$(F(x))^{-\nu}(1 - F^r(x))^{-\mu} = cF'(x)$$

с той же константой c , что и в равенстве (29).

Доказательство теоремы 2. Как и при доказательстве теоремы 1, можем, не умаляя общности, считать для простоты изложения, что константа a в соотношении (28) равна нулю. Используя формулы, аналогичные (19) и (20), можно равенство (28) (при $a = 0$) записать в виде

$$\frac{k-j}{k}x + \frac{j\,(k-j)}{kF^{k-j}\,(x)} \int\limits_0^{F(x)} G\,(y)\,y^{k-j-1}dy \\ + xF^r\,(x) + r\int\limits_{F(x)}^1 G\,(y)\,y^{r-1}dy = \Big(2-\frac{j}{k}\Big)x. \quad (30)$$

Из (30) (после замены $z = F\,(x)$) следует, что

$$\frac{j\,(k-j)}{kz^{k-j}} \int\limits_0^z G\,(y)\,y^{k-j-1}dy + G\,(z)\,z^r + r\int\limits_z^1 G\,(y)\,y^{r-1}dy = G\,(z)$$

и

$$j\,(k-j)\int\limits_0^z G\,(y)\,y^{k-j-1}dy + kz^{k+r-j}G\,(z) \\ + krz^{k-j}\int\limits_z^1 G\,(y)\,y^{r-1}dy = kz^{k-j}G\,(z)\,, \quad 0<z<1. \quad (31)$$

Дважды дифференцируя по z соотношение (31), приходим после соответствующих упрощений к уравнению

$$\Big(\big(k^2-kj+kr+k\big)\,z^r - k - k^2 + 2kj - j^2\Big)\,G'\,(z) \\ = kz\,(1-z^r)\,G''\,(z)\,, \quad 0<z<1. \quad (32)$$

Из (32) следует, что $G'\,(z) = c\,(1-z^r)^\mu\,z^\nu$, $0<z<1$, где

$$\mu = \frac{j^2}{kr} - 1 - \frac{j}{r}, \quad \nu = 2j - \frac{j^2}{k} - k - 1,$$

а c – некоторая положительная константа. Таким образом, $G(x)$ можно записать в виде

$$G(x) = c \int_{1/2}^{x} (1 - z^r)^\mu z^\nu dz + d, \quad 0 < x < 1, \tag{33}$$

где $c > 0$ и $-\infty < d < \infty$ – некоторые константы.

Анализируя поведение $G(x)$ (при x стремящихся к нулю и единице), убеждаемся, что $\gamma = G(0) = -\infty$ и $\delta = G(1) = \infty$. Если в исходном равенстве (28) взять произвольную (не обязательно нулевую) константу a, то мы придем к такому же виду функции $G(x)$, но только в качестве d нужно будет взять другую константу.

Литература

1. M. C. Jones, *Student's simplest distribution.* — The Statistician **51**, Part 1 (2002), 41–49.
2. В. Б. Невзоров, *Об одном свойстве распределения Стьюдента с двумя степенями свободы.* — Зап. научн. семин. ПОМИ **294** (2002), 148–157.
3. В. Б. Невзоров, *Экстремумы и t_2-распределение Стьюдента.* — Зап. научн. семин. ПОМИ **351** (2007), 232–237.
4. V. B. Nevzorov, N. Balakrishnan, M. Ahsanullah, *Simple characterizations of Student's t_2-distribution.* — The Statistician **52** (2003), Part 3, 395–400.
5. В. Б. Невзоров, *О характеризациях распределения Стьюдента с двумя степенями свободы.* — Вестник СПбГУ **1** (2005), вып. 4, 38–42.

Marudova N. M., Nevzorov V. B. On one class of distributions including Student's t_2-distribution.

Some new characterizations of distributions by regressional properties of extremal order statistics are obtained.

С.-Петербургский
государственный университет
E-mail: marudova@gmail.com,
valnev@mail.ru

Поступило 26 ноября 2008 г.