Записки научных
семинаров ПОМИ
Том 294, 2002 г.

В.Б. Невзоров

ОБ ОДНОМ СВОЙСТВЕ РАСПРЕДЕЛЕНИЯ СТЬЮДЕНТА С ДВУМЯ СТЕПЕНЯМИ СВОБОДЫ

1. Введение. В начале прошлого века Уильям Госсет, получивший должность химика в Дублине на пивоваренном производстве Гиннесса, разрабатывал статистические тесты для проверки качества пива. Свои математические работы Госсет подписывал псевдонимом Student. В результате его исследований (см., например, [8]) в статистический обиход вошло новое семейство распределений.

Мы говорим, что случайная величина X имеет t_ν-распределение (или распределение Стьюдента с ν, $\nu > 0$, степенями свободы), если плотность распределения X имеет вид

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

где $\Gamma(s) = \int\limits_0^\infty x^{s-1}\exp(-x)dx$ обозначает гамма-функцию.

Отметим, что если независимые случайные величины Y и Z имеют соответственно стандартное нормальное распределение и χ^2-распределение с n степенями свободы, то дробь $X = \frac{Y\sqrt{n}}{\sqrt{Z}}$ имеет t_n-распределение. Почти во всех учебниках по математической статистике показано, как это свойство распределений Стьюдента с целыми положительными значениями параметра (позволяет использовать t_n-распределения и их процентные точки в статистических задачах, связанных с проверкой гипотез или оцениванием неизвестного математического ожидания a нормального $N(a,\sigma^2)$ распределения в случае, когда параметр σ также является неизвестным.

Совсем недавно Джоунс [4] заинтересовался одним из представителей семейства стьюдентовских распределений, причем

Работа выполнена при поддержке РФФИ (гранты 00-15-96019, 01-01-00031 и 02-01-00779).

этот интерес был связан не со статистическим, а с методологическим использованием этого распределения. Речь идет о t_2-распределении с плотностью

$$p_2(x) = (2 + x^2)^{-3/2} \tag{2}$$

и функцией распределения

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

Джоунс отмечает, что, наряду с t_1-распределением, совпадающим с распределением Коши, распределение (3) является простейшим представителем стьюдентовского семейства. Более того, простая алгебраическая функция в правой части (3) представляется Джоунсу более привлекательной, чем функция распределения Коши

$$F_1(x) = \frac{1}{2} + \frac{1}{\pi}\mathrm{arctgx}\,,$$

выраженная через обратную тригонометрическую функцию. Джоунс предложил использовать t_2-распределение в элементарных вероятностных курсах, поскольку оно позволяет наглядно и просто иллюстрировать различные свойства и характеристики непрерывных распределений. К тем немногим распределениям, которые могли бы в номинации "непрерывные симметричные распределения с бесконечными носителями" конкурировать простотой формы и элементарностью изложения ц t_2-распределением, можно, пожалуй, отнести лишь нормальное, логистическое, а также распределения Лапласа и Коши. В своей статье Джоунс собрал убедительный материал, показывающий, что незаслуженно игнорируемое вероятностным сообществом распределение Стьюдента с двумя степенями свободы достойно стать таким же хорошим учебным примером, как и четыре приведенных выше распределения.

Так получилось, что если Джоунс пришел к распределению (3), рассматривая семейство распределений Стьюдента, то мы нашли это распределение, решая одну характеризационную задачу, совсем в другом семействе, где его родственниками являются такие классические распределения, как логистическое и равномерное. Рассматриваемое нами семейство содержит функции распределения F, удовлетворяющие функциональному уравнению

$$(F(x)(1-F(x))^{\alpha} = cF'(x), \tag{4}$$

где $-\infty < \alpha < \infty$, $c > 0$. Полученные нами результаты показывают, что t_2-распределение может не только служить хорошим учебным примером при изучении непрерывных распределений, но играет также важную роль и в теории порядковых статистик.

2. От логистического к t_2-распределению. Характеризация распределения (3), которую мы приведем ниже, связана со свойствами порядковых статистик.

Пусть X_1, $X_2, \ldots$ – независимые случайные величины, имеющие общую непрерывную функцию распределения, и $X_{1,n} \leqslant \ldots \leqslant X_{n,n}$, $n = 1, 2, \ldots,$ – соответствующие порядковые статистики. В большинстве своем различные свойства порядковых статистик $X_{k,n}$, $1 \leqslant k \leqslant n$, $n = 1, 2, \ldots$, характеризуют (в классе непрерывных распределений) экспоненциальное или равномерное распределение (см., например, обзорные статьи [5] и [7]). Поэтому нас заинтересовал следующий результат ([3]), справедливый для логистического распределения: если

$$F(x) =_F 0\big((x-\mu)/\sigma\big), \quad -\infty < \mu < \infty, \quad \sigma > 0,$$

где

$$F_0(x) = 1/\big(1 + \exp(-x)\big), \quad -\infty < x < \infty,$$

то выборочная середина размаха $M = (X_{1,3} + X_{3,3})/2$ и выборочная медиана $X_{2,3}$ одинаково распределены. До сих пор нет ответа на вопрос, характеризуется ли логистическое распределение свойством одинаковой распределенности случайных величин M и $X_{2,3}$ или существуют другие распределения, для которых это свойство также выполняется.

Мы рассмотрели близкую проблему, решение которой привело к характеризации, но не логистического, а некоторых других распределений, включая t_2. Первоначально мы задались целью выяснить, для каких распределений выполняется равенство

$$\varphi(x) = x \quad \text{п.н.} \tag{5}$$

для регрессионной функции $\varphi(x) = E(M|X_{2,3} = x)$. Выяснилось, что (5) справедливо тогда и только тогда, когда

$$F(x) = F2\big((x-\mu)/\sigma\big), \quad -\infty < \mu < \infty, \quad \sigma > 0,$$

где $F_2(x)$ определена соотношением (3). Естественным обобщением соотношения (5) явилось равенство

$$\varphi_k(x) = x \quad \text{п.н.} \tag{6}$$

для регрессионных функций

$$\varphi_k(x) = E\big((X_{k,2k+1} + X_{k+2,2k+1})/2 \,|\, X_{k+1,2k+1} = x\big), \quad k = 1, 2, \dots . \tag{7}$$

В определении (7) фигурируют выборочные медианы $X_{k+1,2k+1}$ и выборочные середины квази-размахов $(X_{k,2k+1} + X_{k+2,2k+1})/2$. Удалось найти семейство распределений, характеризуемых при различных значениях к соотношениями (6). Это семейство, существенно отличающееся от стьюдентовского, естественно содержит и t_2-распределение.

3. Характеризация t_2-распределения. Мы выделим в отдельный параграф формулировку и доказательство результата для t_2-распределения.

Теорема 1. *Пусть X_1, X_2 и X_3 – независимые одинаково распределенные случайные величины с конечными математическими ожиданиями, функцией распределения $F(x)$ и плотностью $f(x)$. Пусть также $f(x) > 0$, если $\gamma < x < \delta$, где $\gamma = \inf x : F(x) > 0$ и $\delta = \sup x : F(x) < 1$. Тогда эквивалентны следующие утверждения:*

a) $E(M|X_{2,3} = x) = x$ *п.н.;* (8)

b) $E(\overline{X}|X_{2,3} = x) = x$ *п.н.,*
где $\overline{X} = (X_1 + X_2 + X_3)/3$ *является выборочным средним;* (9)

c) $E(X_1 \,|\, X_{2,3} = x) = x$ *п.н.;* (10)

d) $(F(x)(1 - F(x))^{3/2} = cf(x)$, $\gamma < x < \delta$, $c > 0$; (11)

e) $\gamma = -\infty$, $\delta = \infty$ *и* $F(x) = F_2((x - \mu)/\sigma)$, $-\infty < \mu < \infty$, $\sigma > 0$, *где* F_2 *определена в* (3).

Доказательство теоремы 1. Эквивалентность соотношений (8)–(10) очевидна. Покажем теперь, что из (10) следует (11). Известно (см., например, [6]), что

$$E(X_1|X_{k,n} = x) = \frac{x}{n} + \frac{k-1}{n} E(X|X < x) + \frac{n-k}{n} E(X|X > x) \quad \text{п.н.} \tag{12}$$

В нашем случае приходим к равенству

$$x = E(X_1 \mid X_{2,3} = x) = \\ \frac{1}{3}\left(x + \frac{1}{F(x)} \int\limits_{-\infty}^{x} tf(t)dt + \frac{1}{1-F(x)} \int\limits_{x}^{\infty} tf(t)dt\right). \quad (13)$$

В условиях теоремы $E|X_1| < \infty$. Следовательно, справедливы равенства

$$\lim_{x\to-\infty} xF(x) = \lim_{x\to-\infty} x(1-F(x)) = 0,$$

$$\frac{1}{F(x)} \int\limits_{-\infty}^{x} tf(t)dt = x - \frac{1}{F(x)} \int\limits_{-\infty}^{x} F(t)dt \quad (14)$$

и

$$\frac{1}{1-F(x)} \int\limits_{x}^{\infty} tf(t)dt) = x + \frac{1}{1-F(x)} \int\limits_{x}^{\infty} (1-F(t))dt. \quad (15)$$

Из соотношений (13)–(15) получаем, что (10) равносильно равенству

$$F(x) \int\limits_{x}^{\infty} (1-F(t))dt = (1-F(x)) \int\limits_{-\infty}^{x} F(t)dt, \quad \gamma < x < \delta, \quad (16)$$

которое можно переписать в виде

$$R'(x) = 0, \quad \gamma < x < \delta,$$

где

$$R(x) \int\limits_{0}^{x} = F(t)dt \int\limits_{x}^{\infty} (1-F(t))dt.$$

Следовательно, для всех $x \in (\gamma, \delta)$ справедливо равенство

$$\int\limits_{-\infty}^{x} F(t)dt \int\limits_{x}^{\infty} (1-F(t))dt = c, \quad (17)$$

где c здесь и далее обозначает положительные константы (не обязательно одинаковые в разных соотношениях). Последовательно избавляясь с помощью дифференцирования от интегралов в левой части (17), получаем, что

$$\big(F(x)(1-F(x))\big)^{3/2} = cf(x), \quad \gamma < x < \delta,$$

т.е. выполняется равенство (11).

Пусть теперь $G(x)$ обозначает функцию, обратную $F(x)$. В условиях теоремы G непрерывна и строго монотонна для $0 < x < 1$,

$$G(0) = \lim_{x\to 0} G(x) = \gamma$$

и

$$G(1) = \lim_{x\to 1} G(x) = \delta.$$

Все это позволяет переписать (11) в виде

$$cf(G(x)) = \big(x(1-x)\big)^{3/2}, \quad 0 < x < 1. \tag{18}$$

Поскольку $G'(x) = 1/f(G(x))$, из (18) следует, что

$$G'(x) = s/\big(x(1-x)\big)^{3/2},$$

и

$$G(x) = c\int_{1/2}^{x} \frac{du}{(u(1-u))^{3/2}} + d, \quad 0 < x < 1, \tag{19}$$

где $s > 0$ и $-\infty < d < \infty$. Из (19) вытекает, что $\gamma = G(0) = -\infty$, $\delta = G(1) = \infty$ и

$$G(x) = cG_1(x) + d,$$

где

$$G_1(x) = \int_{1/2}^{x} \frac{du}{(u(1-u))^{3/2}} = \frac{2x-1}{2\sqrt{x-x^2}}, \quad 0 < x < 1, \tag{20}$$

представляет собой функцию, обратную функции распределения

$$H(x) = \frac{1}{2}\bigg(1 + \frac{x}{\sqrt{1+x^2}}, \quad -\infty < x < \infty,$$

а тогда $F(x)$ – функция, обратная $G(x)$, находится из равенства

$$F(x) = H((x-d)/c) = \frac{1}{2}\bigg(1 + \frac{x-d}{\sqrt{c^2+(x-d)^2}}\bigg), \qquad c > 0, \quad -\infty < d < \infty. \tag{21}$$

Заметим, что $H(x)$ и функция $F2(x)$, определенная в (3), принадлежат одному типу распределений. Следовательно, семейство функций распределения (21) совпадает с семейством

$$F(x) = F_2((x-\mu)/\sigma), \quad -\infty < \mu < \infty, \quad \sigma > 0. \tag{22}$$

Мы показали, что если выполняется соотношение (10), то функция распределения F обязана принадлежать семейству (22). Для завершения доказательства несложно убедиться, что для любой функции (21) справедливо равенство (16), а следовательно, имеют место утверждения а), b) и s) теоремы.

4. Обобщение теоремы 1. Утверждения теоремы 1 допускают различные обобщения. Мы ограничимся рассмотрением равенств, аналогичных (8), для произвольной выборки нечетного объема.

Пусть $X_1, X_2, \ldots$ – независимые слу чайные величины с конечными математическими ожиданиями, функцией распределения $F(x)$ и плотностью $f(x)$, а $X_{1,n} \leqslant \ldots \leqslant X_{n,n}$, $n = 1, 2, \ldots$, – соответствующие порядковые статистики. Будем предполагать, что $f(x) > 0$, если $\gamma < x < \delta$, где $\gamma = \inf x : F(x) > 0$ и $\delta = \sup x : F(x) < 1$. Как и в теореме 1, $G(x)$ обозначает функцию, обратную $F(x)$. Для любого $k = 1, 2, \ldots$ рассмотрим выборочную середину квази-размаха

$$M_k = (X_{k,2k+1} + X_{k+2,2k+1})/2,$$

которая при $k = 1$ совпадает со случайной величиной M из теоремы 1. Справедлив следующий результат.

Теорема 2. *Для любого $k = 1, 2, \ldots$ эквивалентны утверждения*

a) $E(M_k \mid X_{k+1,2k+1} = x) = x$, $\gamma < x < \delta$, *п.н.;* (23)

b) $(F(x)(1-F(x))^{k/2+1} = cf(x)$, $\gamma < x < \delta$, (24)
где s – некоторая положительная константа;

c) $\gamma = -\infty$, $\delta = \infty$

$$G(x) = c \int_{1/2}^{x} \big(u(1-u)\big)^{-(k/2+1)} du + d, \quad 0 < x < 1, \tag{25}$$

где $s > 0$ *и* $-\infty < \infty$.

Доказательство теоремы 2. Условное распределение вектора $(X_{1,2k+1}, \dots, X_{k,2k+1})$ при условии, что $X_{k+1,2k+1} = x$, совпадает с безусловным распределением порядковых статистик $Y_{1,k}, \dots, Y_{k,k}$, построенных по независимым случайным величинам $Y_1, \dots, Y_k$ с общей функцией распределения $F_x(u) = F(u)/F(x)$, $u < x$ (см., например, теорему 4.1. в [1]). Поэтому

$$E\big(X_{k,2k+1} \,|\, X_{k+1,2k+1} = x\big) = EY_{k,k} = \int\limits_{-\infty}^{x} ud\big((Fx(u))^k\big) =$$

$$xF_x^k(x) - \int\limits_{-\infty}^{x} F^k(u)du = x - \frac{1}{F^k(x)} \int\limits_{-\infty}^{x} F^k(u)du. \qquad (26)$$

Используя аналогичный результат для вектора $(X_{k+2,2k+1}, \dots, X_{2k+1,2k+1})$, получаем, что

$$E(X_{k+2,2k+1} \,|\, X_{k+1,2k+1} = x) = x + \frac{1}{(1-F(x))^k}(1-F(u))^k du. \quad (27)$$

Из (26) и (27) следует более общее, чем (16), равенство

$$F^k(x)\int\limits_{x}^{\infty} \big(1-F(u)\big)^k du = \big(1-F(u)\big)^k F^k(u)du, \quad \gamma < x < \delta, \qquad (28)$$

из которого получаем соотношение

$$\int\limits_{-\infty}^{x} F^k(u)du \int\limits_{x}^{\infty} \big(1-F(u)\big)^k du = s \qquad (29)$$

с некоторой положительной константой s. После соответствующего дифференцирования мы приходим к равенству (24), а затем уже методом, знакомым нам по теореме 1, получаем более общее, чем (19), соотношение (25).

Замечание 1. Из утверждений теоремы 2 следует, что свойство (23) порядковых статистик характеризует некоторые функции распределения из семейства распределений, задаваемых уравнениями вида

$$\big(F(x)(1-F(x)\big)^{\alpha} = cF'((x), \quad c > 0, \quad -\infty < x < \infty. \qquad (30)$$

Распределения, удовлетворяющие (23), отвечают значениям $\alpha = k/2$ при $k = 3, 4, \ldots$. В частности, распределения стьюдентовского t_2-типа соответствуют $\alpha = 3/2$. Интересно отметить, что семейство (30) содержит также такие типы классических распределений, как логистический ($\alpha = 1$) и равномерный ($\alpha = 0$). При $\alpha = 1/2$ равенство (30) характеризует функции распределения

$$F(x) = F_{1/2}\big((x-d)/c\big), \quad c > 0, \quad -\infty < d < \infty,$$

где

$$F_{1/2}(x) = \sin^2 x, \quad 0 \leqslant x \leqslant \pi/2.$$

Более подробную информацию о некоторых других распределениях, удовлетворяющих (30), можно найти в работе [5].

Замечание 2. В нашем введении мы упоминали пожелание Джоунса присоединить t_2 – распределение к группе из четырех классических непрерывных распределений с бесконечными носителями (логистическое, нормальное, Лапласа, Коши), изучаемых в большинстве вероятностных курсов. Мы можем добавить, что такое присоединение объединит четыре (на первый взгляд, не связанных друг с другом) указанных типа распределений, принимая во внимание их различные родственные отношения к t_2 – распределению. Действительно, логистическое и t_2 распределение принадлежат общему семейству (30), распределение Коши (оно же t_1) и t_2 представляют стьюдентовское семейство. Наконец, если возьмем независимые случайные величины Y и Z, имеющие соответственно стандартное нормальное распределение и распределение Лапласа с плотностью $\exp(-|x|)/2$, то отношение $Y/|Z|^{1/2}$ даст нам случайную величину с t_2 – распределением.

Литература

1. В. Б. Невзоров, *Рекорды. Математическая теория.* ФАЗИС, М. (2000).
2. U. Gather, U. Kamps, N. Schweitzer, *Characterizations of distributions via identically distributed functions of order statistics.* — In: Handbook of statistics-16 (Eds. N. Balakrishnan, C. R.Rao), Elsevier, Amsterdam (1998), pp. 257–290.
3. E. O. George, C. C. Rousseau, *On the logistic midrange.* — Ann. Inst. Statist. Math. **39** (1987), 627–635.
4. M. C. Jones, *Student's simplest distribution.* — The Statistician **51**, Part 1 (2002), 41–49.
5. U. Kamps, *A general recurrence relation for moments of order statistics in a class of probability distributions and characterizations.* — Metrika **38** (1991), 215–225.

6. H. N. Nagaraja, V. B. Nevzorov, *On characterizations based on record values and order statistics.* — J. Stat. Plann. and Infer. **63** (1997), 271–284.
7. C. R. Rao, D. N. Shanbhag,, *Recent approaches to characterizations based on order statistics and record values.* — In: Handbook of statistics-16 (Eds.: N. Balakrishnan, C. R. Rao), Elsevier, Amsterdam (1998), pp. 231–256.
8. "Student" (W. Gosset), *The probable error of a mean.* — Biometrika **6** (1908), 1–25.

Nevzorov V. B. One property of Student's distribution on two degrees of freedom.

Characterizations of Student's t_2-distribution and some related distributions based on regressional properties of order statistics are obtained.

С.-Петербургский
государственный университет

Поступило 21 октября 2002 г.