Записки научных
семинаров ПОМИ
Том 341, 2007 г.

В. К. Сагателян

ХАРАКТЕРИЗАЦИЯ РАСПРЕДЕЛЕНИЙ РАВЕНСТВОМ ПОРЯДКОВЫХ СТАТИСТИК

1. Введение

Пусть $X_1, X_2, \ldots$ – независимые случайные величины, имеющие общую функцию распределения $F(x)$ и пусть для любого $n = 1, 2, \ldots,$

$$X_{1,n} \leqslant X_{2,n} \leqslant \ldots \leqslant X_{n,n}$$

обозначают порядковые статистики, соответствующие случайным величинам $X_1, X_2, \ldots, X_n$. Теория порядковых статистик рассматривается в монографиях [1]–[3]. Существует множество характеризаций функций распределения $F(x)$, основанных на разных свойствах порядковых статистик и их моментных характеристик. Ниже мы приведем еще один результат, в котором равенство порядковых статистик характеризует двухточечные распределения.

2. Характеризация распределений

Рассмотрим последовательность независимых случайных величин $X_1, X_2, \ldots$, имеющих общую функцию распределения $F(x)$, и порядковые статистики $X_{1,n} \leqslant X_{2,n} \leqslant \ldots \leqslant X_{n,n}$ и $X_{1,m} \leqslant X_{2,m} \leqslant \ldots \leqslant X_{m,m}$. Существуют ли такие невырожденные распределения и такие n, m, k, l, для которых выполняется равенство

$$X_{k,n} \stackrel{d}{=} X_{l,m}, \tag{1}$$

тоесть равенство по распределению? Справедлив следующий результат.

Теорема. *Если выполняется равенство $X_{k,n} \stackrel{d}{=} X_{l,m}$, $m \neq n$, то исходная функция распределения является либо вырожденной, либо имеет лишь две точки роста.*

Замечание 1. Не умаляя общности, можно предположить, что $n > m$. Если $n = m$, то очевидно, что равенство (1) имеет место только при $k = l$.

Действительно, функции распределения порядковых статистик $X_{k,n}$ и $X_{l,m}$ имеют следующий вид:

$$F_{k,n}(x) = \sum_{r=k}^{n} C_n^r F^r(x)(1 - F(x))^{n-r},$$

$$F_{l,m}(x) = \sum_{r=l}^{m} C_m^r F^r(x)(1 - F(x))^{m-r}$$

соответственно. Через $R(x)$ обозначим разность этих двух функций. Тогда $R(x) = \sum\limits_{r=k}^{n} C_n^r F^r(x)\ (1 - F(x))^{n-r} - \sum\limits_{r=l}^{m} C_m^r F^r(x)(1 - F(x))^{m-r}$. Если $n = m$ и выполнено (1), приходим к равенству

$$\sum_{r=k}^{n} C_n^r F^r(x)(1 - F(x))^{n-r} - \sum_{r=l}^{n} C_n^r F^r(x)(1 - F(x))^{n-r} = 0.$$

Пусть $k \neq l$, для определенности возьмем $k > l$. Тогда

$$R(x) = -\sum_{r=l}^{k-1} C_n^r F^r(x)(1 - F(x))^{n-r} = 0,$$

т.е. $F(x) = 0$ или $F(x) = 1$ для любого x, что возможно только для вырожденных распределений.

Замечание 2. Из доказательства теоремы будет видно, что исходная функция распределения является двухточечной в том случае, когда индексы порядковых статистик в равенстве (1) удовлетворяют условиям $n - m > k - l > 0$, или $m - n > l - k > 0$.

Доказательство теоремы. Не умаляя общности, полагаем, что $n > m$. Обозначим $F(x) = z$. Тогда равенство $F_{k,n}(x) = F_{l,m}(x)$ будет равносильно соотношению

$$\sum_{r=k}^{n} C_n^r z^r (1 - z)^{n-r} = \sum_{r=l}^{m} C_m^r z^r (1 - z)^{m-r}.$$

Поскольку полином n-ой степени имеет не более n корней, получаем, что $F(x)$-дискретная функция распределения, и т.к. значения $z = 0$, $z = 1$ заведомо являются корнями, то $F(x)$ может иметь не более $n - 1$ точек роста. Рассмотрим функцию

$H(z) = F_1(z) - F_2(z)$, где $F_1(z) = \sum\limits_{r=k}^{n} C_n^r z^r (1-z)^{n-r}$, $F_2(z) = \sum\limits_{r=l}^{m} C_m^r z^r (1-z)^{m-r}$, $0 < z < 1$, – функции распределения равномерных порядковых статистик $U_{k,n}$ и $U_{l,m}$ соответственно. Видим, что $R(x) = H(F(x))$. Справедливы равенства

$$F_1(z) = \sum_{r=k}^{n} C_n^r z^r (1-z)^{n-r} = \int_0^z \frac{n!}{(k-1)!(n-k)!} t^{k-1}(1-t)^{n-k} dt,$$

$$F_2(z) = \sum_{r=l}^{n} C_m^r z^r (1-z)^{m-r} = \int_0^z \frac{m!}{(l-1)!(m-l)!} t^{l-1}(1-t)^{m-l} dt.$$

Очевидно, что $H(0) = 0, H(1) = 0$, требуется найти остальные корни уравнения $H(z) = 0$. Рассмотрим производную $g(z) = H'(z)$. Имеем

$$g(z) = \frac{n!}{(k-1)!(n-k)!} z^{k-1}(1-z)^{n-k} - \frac{m!}{(l-1)!(m-l)!} z^{l-1}(1-z)^{m-l} =$$

$$= kC_n^k z^{l-1}(1-z)^{m-l}[z^{k-l}(1-z)^{(n-k)-(m-l)} - \frac{lC_m^l}{kC_n^k}].$$

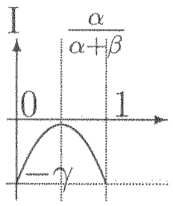
Обозначим $\frac{lC_m^l}{kC_n^k} = \gamma$. Получаем, что если $k > l$, то $g(0) = \begin{cases} \{0, l > 1 \\ -C_m^1, l = 1 \end{cases}$. Аналогично получаем, что если $n - k > m - l$, то $g(1) = \begin{cases} \{0, m > l \\ -mC_m^m, m = l \end{cases}$. Рассмотрим случай $l > 1, m > l$ и функцию

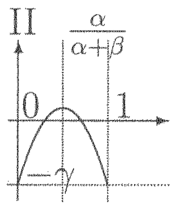
$$g_{\alpha,\beta}(x) = x^\alpha (1-x)^\beta - \gamma,$$

где $\alpha = k - l, \beta = (n-k) - (m-l)$. В нашей ситуации $g_{\alpha,\beta}(0) = -\gamma$ и $g_{\alpha,\beta}(1) = -\gamma$. Производная имеет вид

$$g'_{\alpha,\beta}(x) = \alpha x^{\alpha-1}(1-x)^\beta - \beta(1-x)^{\beta-1}x^\alpha = x^{\alpha-1}(1-x)^{\beta-1}(\alpha - (\alpha+\beta)x).$$

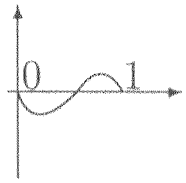
Следовательно, $g'_{\alpha,\beta}(x) > 0$, если $0 < x < \frac{\alpha}{\alpha+\beta}$ и $g'_{\alpha,\beta}(x) < 0$, если $\frac{\alpha}{\alpha+\beta} < x < 1$. Таким образом, функция $g_{\alpha,\beta}(x)$ возрастает в области $0 < x < \frac{\alpha}{\alpha+\beta}$ и убывает в области $\frac{\alpha}{\alpha+\beta} < x < 1$. Возникают два случая:





т.е. у функции $g_{\alpha,\beta}(x)$ могут быть максимум два корня.

Функция $g(z)$ имеет вид $g(z) = kC_n^k z^{l-1}(1-z)^{m-l} g_{\alpha,\beta}(z)$ и является производной $H(z)$. Следовательно, там где $g(z) < 0$, $H(z)$ убывает, а там где $g(z) > 0$, $H(z)$ возрастает. Поскольку $H(0) = H(1) = 0$, то мы можем иметь дело только со вторым случаем, когда функция $H(z)$ вначале убывает, затем возрастает и снова убывает.



Поэтому, в лучшем случае $H(z)$ кроме $z = 0$ и $z = 1$ может иметь еще только один корень, т.е. исходное распределение имеет не более двух точек роста.

Опишем поведение функции $H(z)$ в окрестности нуля и единицы, в том случае, когда $n - k > m - l > 0$. Итак,

$$H(z) = kC_n^k \int\limits_0^z t^{k-1}(1-t)^{n-k} dt - lC_m^l \int\limits_0^z t^{l-1}(1-t)^{m-l} dt,$$

и в силу того, что $k > l$, получаем, что $H(z) \sim C_m^l z^l$, при $z \to 0$. Следовательно, $H(0+0) < 0$. Аналогично получаем, что $H(z) \sim lC_m^l \frac{(1-z)^{m-l+1}}{(m-l+1)}$, при $z \to 1$. Таким образом $H(1-0) > 0$.

Итак, для случая $n - m > k - l > 0$ у функции $H(z)$ кроме $z = 0$ и $z = 1$ существует еще одна точка $0 < z < 1$, где $H(z) = 0$. Доводы о поведении $g(z)$, а следовательно и $H(z)$ были справедливы лишь если $n - m > k - l$ и $m > l > 1$. Аналогично рассматриваются остальные случаи. Теорема доказана.

3. Примеры двухточечных распределений, характеризуемых равенством порядковых статистик

Пример 1. Обозначим через p именно те значения (отличные от нуля и единицы) функций $F(x)$, при которых выполняются разные частные случаи равенства (1).

Рассмотрим равенство $X_{m+1,m+2} \stackrel{d}{=} X_{m,m}$, обозначим $z = F(x)$. Решая уравнение $H(z) = 0$, где

$$H(z) = \sum_{r=m+1}^{m+2} C_{m+2}^r z^r (1-z)^{m+2-r} - \sum_{r=m}^{m} C_m^r z^r (1-z)^{m-r} =$$

$$= (m+2)z^{m+1} - (m+1)z^{m+2} - z^m,$$

получаем, что $z = \frac{1}{m+1}$ и соответственно $p = \frac{1}{m+1}$.

Далее, действуя аналогично, для случая $X_{m,m+2} \stackrel{d}{=} X_{m-1,m}$, получим значение $p = \frac{2}{m+1}$.

Пример 2. Рассмотрим равенство $X_{m-l+1,m+1} \stackrel{d}{=} X_{m-l,m-1}$. Убедимся, что данном случае $p = \frac{l}{m}$. Как и прежде, из соотношения $F_{m-l+1,m+1}(x) = F_{m-l,m-1}(x)$, обозначив $z = F(x)$, получим уравнение следующего вида

$$\sum_{r=m-l+1}^{m+1} C_{m+1}^r z^r (1-z)^{m+1-r} - \sum_{r=m-l}^{m-1} C_{m-1}^r z^r (1-z)^{m-1-r} = 0. \quad (2)$$

Первую сумму в левой части уравнения обозначим через $H_1(z)$, а вторую $H_2(z)$. Функции $H_1(z)$ и $H_2(z)$ являются функциями распределения равномерных порядковых статистик $U_{m-l+1,m+1}$ и $U_{m-l,m-1}$ соответственно. Отсюда

$$H_1(z) = (m+1)C_m^l \int_0^z t^{m-1}(1-t)^l dt$$

и

$$H_2(z) = (m-1)C_{m-2}^{l-1} \int_0^z t^{m-l-1}(1-t)^{l-1} dt,$$

а уравнение (2) принимает вид

$$H_1(z) - H_2(z) = (m+1)C_m^l \int\limits_0^z t^{m-l}(1-t)^l dt -$$

$$-(m-1)C_{m-2}^{l-1} \int\limits_0^z t^{m-l-1}(1-t)^{l-1} dt = 0. \qquad (3)$$

Остается показать, что

$$H_1(z) - H_2(z) = \frac{(m-1)!}{(m-l)!l!}(1-z)^l z^{m-l}(mz-l), \qquad (4)$$

а тогда, поскольку $z = \frac{l}{m}$ является корнем этого уравнения, это же z будет также корнем уравнения (3), а значит и (2). Нетрудно заметить, что при $z = 0$ выражения (4) и (3) совпадают. Остается показать, что совпадают также и производные выражений (4) и (3). Для этого достаточно просто продифференцировать эти выражения.

Литература

1. В. Б. Невзоров, *Рекорды. Математическая теория.* Фазис, М. (2000).
2. M. Ahsanullah, V. B. Nevzorov, *Ordered Random Variables.* NSP, NY. (2001).
3. B. C. Arnold, N. Balakrishnan, H. N. Nagaraja, *A first course in order statistics.* Wiley, NY. (1992).

Saghatelyan V. K. Characterization of distributions by equalities of order statistics.

It is proved that the equality of two different order statistics holds for degenerate or two-point distributions only. Several samples of such equalities are considered and particular values of corresponding distribution functions are obtained.

Санкт-Петербургский государственный университет

E-mail: vahagn_s@yahoo.com

Поступило 20 декабря 2006 г.