Записки научных
семинаров ПОМИ
Том 368, 2009 г.

А. Берред, В. Б. Невзоров

ВЫБОРКИ БЕЗ ВОЗВРАЩЕНИЯ: ОБ ОДНОМ СВОЙСТВЕ РАСПРЕДЕЛЕНИЯ РАЗМАХА

Рассмотрим некоторую конечную совокупность $\Pi = \{x_1 < x_2 < \cdots < x_N\}$, состоящую из N различных элементов, случайную выборку без возвращения $Y_1, Y_2, \dots, Y_n$ из этой совокупности, где $n \leqslant N$, и соответствующие порядковые статистики $Y_{1,n} < Y_{2,n} < \cdots < Y_{n,n}$. Отметим, что при случайном выборе без возвращения элементы совокупности Π отбираются по очереди, отобранные элементы в дальнейшем выборе не участвуют, для всех объектов, еще не отобранных на данный момент, вероятность быть выбранными на следующем этапе одинакова.

В работе [1] был приведен ряд соотношений для различных распределений, связанных с порядковыми статистиками $Y_{k,n}$. В частности, было показано, что при любом $k = 2, 3, \dots, n-1$ случайные векторы

$$Y^{(1)} = (Y_{1,n}, \dots, Y_{k-1,n}) \quad \text{и} \quad Y^{(2)} = (Y_{k+1,n}, \dots, Y_{n,n})$$

условно независимы при любом фиксированном значении порядковой статистики $Y_{k,n}$. Более того, условное распределение вектора $Y^{(1)}$ при условии, что $Y_{k,n} = x_{\alpha(k)}$, где $k \leqslant \alpha(k) \leqslant N - n + k$, совпадает с распределением вектора порядковых статистик $(V_{1,k-1}, \dots, V_{k-1,k-1})$, соответствующего выборке без возвращения $V_1, V_2, \dots, V_{k-1}$ из совокупности $\{x_1 < \cdots < x_{\alpha(k)-1}\}$. Аналогично, условное распределение вектора $Y^{(2)}$ (при том же условии) совпадает с распределением вектора $(W_{1,n-k}, \dots, W_{n-k,n-k})$, построенного по выборке без возвращения $W_1, \dots, W_{n-k}$ из совокупности $\{x_{\alpha(k)+1}, \dots, x_N\}$. Данные свойства случайных величин $Y_{1,n}, Y_{2,n}, \dots, Y_{n,n}$ позволили получить (см., например, работы [1] и [2]) некоторые характеризации распределений, основанные на различных регрессионных соотношениях, связывающих эти порядковые статистики. В частности, при $n = 3$ рассма-

Ключевые слова : порядковые статистики, выборки без возвращения, характеризации распределений.

Работа частично поддержана грантами РФФИ 07-01-00688 и 09-01-00808.

тривалась середина размаха $M_3 = (Y_{1,3}+Y_{3,3})/2$ и было показано, что соотношение

$$E(M_3 \mid Y_{2,3}) = Y_{2,3} \tag{1}$$

выполняется для тех и только тех исходных совокупностей

$$\Pi = \{x_1 < x_2 < \cdots < x_N\}, \quad N \geqslant 3,$$

элементы которых можно представить в виде $x_k = cy_k + d$, $k = 1, 2, \ldots, N$, где $c > 0$ и $-\infty < d < \infty$ – произвольные константы, а $y_1 < y_2 < \cdots < y_N$ определяются равенствами

$$y_k = -y_{N-k+1} = \sqrt{N}\frac{(2k-N-1)\Gamma(N-1)\Gamma(k-\frac{1}{2})\Gamma(N-k+\frac{1}{2})}{\sqrt{2\Gamma}\,(k)\Gamma(1/2)\Gamma(N-\frac{1}{2})\Gamma(N-k+1)},$$
$$k = \left[\frac{N}{2}+1\right], \ldots, N. \tag{2}$$

Если рассмотреть последовательность случайных величин Y_N, принимающих с равными вероятностями значения $y_1, y_2, \ldots, y_N$, определяемые равенствами (2), то выясняется , что при $N \to \infty$ справедливо предельное соотношение

$$P\{Y_N < x\} \to \frac{1}{2}\Big(1 + \frac{x}{\sqrt{1+x^2}}\Big), \quad -\infty < x < \infty, \tag{3}$$

т.е. последовательность случайных величин Y_N имеет в пределе распределение Стьюдента с двумя степенями свободы.

При оценивании параметров сдвига и масштаба в паре с серединой размаха $M_n = (Y_{1,n}+Y_{n,n})/2$ часто используется размах выборки $W_n = (Y_{n,n}-Y_{1,n})$. Ниже мы рассмотрим некоторое регрессионное соотношение, близкое к соотношению (1), в котором будет участвовать размах выборки $W_3 = (Y_{3,3} - Y_{1,3})$.

В терминах спейсингов (приращений) порядковых статистик $Y_{3,3} - Y_{2,3}$ и $Y_{2,3} - Y_{1,3}$ равенство (1) можно переписать в виде

$$E(Y_{3,3} - Y_{2,3} \mid Y_{2,3}) - E(Y_{2,3} - Y_{1,3} \mid Y_{2,3}) = 0. \tag{4}$$

Рассмотрим подобное соотношение для сумм этих спейсингов:

$$E(Y_{3,3} - Y_{2,3} \mid Y_{2,3}) + E(Y_{2,3} - Y_{1,3} \mid Y_{2,3}) = C, \tag{5}$$

где $C > 0$ – некоторая константа. Это соотношение эквивалентно равенству

$$E(W_3 \mid Y_{2,3}) = C. \tag{6}$$

Нам нужно найти все совокупности $\Pi = \{x_1 < x_2 < \cdots < x_N\}$, для которых справедливо (6) с некоторой положительной константой C в правой части.

Отметим, что соотношение (6) не меняется при сдвиге элементов совокупности Π на одну и ту же величину, т.е., если оно справедливо для некоторого набора $x_1 < x_2 < \cdots < x_N$, то не изменится при переходе к величинам $a + x_1 < a + x_2 < \cdots < a + x_N$ при любом a. В то же время, если (6) справедливо для совокупности Π с некоторой определенной константой C, то масштабные изменения, т.е. переход к совокупности

$$b\Pi = \{bx_1 < bx_2 < \cdots < bx_N\},$$

где $b > 0$, обеспечивают выполнение (6) уже с константой bC вместо C. Поэтому достаточно с самого начала найти все варианты последовательностей $\Pi = \{x_1 < x_2 < \cdots < x_N\}$, для которых справедливо соотношение

$$E(Y_{3,3} - Y_{1,3} \mid Y_{2,3}) = 1, \tag{7}$$

чтобы затем, переходя к совокупностям вида $C\Pi + a$, получить все последовательности $\{x_1 < x_2 < \cdots < x_N\}$, которые удовлетворяют равенству (6).

Рассмотрим равенство (7). Из соотношений (13) и (14) работы [1] следует, что

$$E(Y_{3,3} \mid Y_{2,3} = x + k) = (x_{k+1} + \cdots + x_N)/(N - k), \tag{8}$$

и

$$E(Y_{1,3} \mid Y_{2,3} = x_k) = (x_1 + \cdots + x_{k-1})/(k - 1) \tag{9}$$

для всех x_k, $k = 2, 3, \ldots, N - 1$. Равенства (7)–(9) приводят к системе уравнений

$$\frac{x_{k+1} + x_{k+2} + \cdots + x_N}{N - k} - \frac{x_1 + x_2 + \cdots + x_{k-1}}{k - 1} = 1, \qquad k = 2, 3, \ldots, N - 1. \tag{10}$$

Сделав замену

$$z_k = x_k - 2k/(N + 1), \quad 1 \leqslant k \leqslant N, \tag{11}$$

приходим к равенствам

$$\frac{z_{k+1}+z_{k+2}+\cdots+z_N}{N-k}-\frac{z_1+z_2+\cdots+z_{k-1}}{k-1}=0,$$
$$k=2,3,\dots,N-1. \quad (12)$$

Соотношения (12) не меняются, если все величины $z_1, z_2, \dots, z_N$ сдвинуть одновременно на одну и ту же величину. Поэтому, не умаляя общности , можем считать, что кроме $(N-2)$ равенств (12) справедливо также соотношение

$$z_1+z_2+\cdots+z_N=0. \quad (13)$$

Из (12) и (13) нетрудно уже получить, что

$$z_k=-(N-1)(z_1+z_2+\cdots+z_{k-1})/(k-1), \quad k=2,3,\dots,N,$$

и тогда

$$(z_1+z_2+\cdots+z_k)=\frac{(N-k)(z_1+z_2+\cdots+z_{k-1})}{k-1},$$
$$k=2,3,\dots,N. \quad (14)$$

Из равенств (14) следует, что

$$(z_1+z_2+\cdots+z_k)=(-1)^{k-1}\frac{(N-2)!}{(N-k-1)!(k-1)!}z_1$$

и

$$z_k=(-1)^{k-1}\frac{(N-1)!}{(N-k)!(k-1)!}z_1=(-1)^{k-1}\mathbf{C}_{N-1}^{k-1}z_1,$$
$$k=2,3,\dots,N. \quad (15)$$

Возвращаясь к переменным $x_1, x_2, \dots, x_N$, получаем, что

$$x_k=\frac{2k}{N+1}+(-1)^{k-1}\mathbf{C}_{N-1}^{k-1}z_1, \quad k=1,2\dots,N, \quad (16)$$

а с учетом того, что соотношение (7) остается в силе при одновременном сдвиге элементов совокупности П, получаем, что оно будет справедливо, если $x_1, x_2, \dots, x_N$ имеют вид

$$x_k=a+\frac{2k}{N+1}+(-1)^{k-1}\mathbf{C}_{N-1}^{k-1}b, \quad k=1,2,\dots,N, \quad (17)$$

где $-\infty < a < \infty$, а в качестве b можно взять любую константу, обеспечивающую монотонность последовательности $x_1 < x_2 < \cdots < x_N$.

Пример. Пусть $N = 5$. Тогда соответствующие равенства (17) принимают вид:

$$x_1 = a + b + 1/3; \qquad x_2 = a - 4b + 2/3; \quad x_3 = a + 1 + 6b;$$
$$x_4 = a - 4b + 4/3; \qquad x_5 = a + b + 5/3.$$

Из этих соотношений следует, что $x_1 < x_2 < \cdots < x_5$, только если $-1/30 < b < 1/30$.

Замечание 1. Если мы аналогично вместо (7) рассмотрим соотношение (6), то соответствующее решение можно будет записать в виде:

$$y_k = Cx_k, \quad k = 1, 2, \ldots, N, \tag{18}$$

где C – положительная константа из правой части (6), а $x_1 < x_2 < \cdots < x_N$ – любая последовательность, определенная равенствами (17).

Замечание 2. Несложные рассуждения показывают, что если при $N \to \infty$ рассмотреть последовательность совокупностей элементов $\Pi_N = \{x_1 < x_2 < \cdots < x_N\}$, для которых соотношение (6) справедливо, то можно убедиться, что в пределе соответствующие последовательности дискретных распределений могут сходиться лишь к одному из равномерных на интервалах вида $(a - C, a + C)$ распределений, где $-\infty < a < \infty$, а C – положительная константа из правой части (6).

Литература

1. А. Берред, В. Б. Невзоров, *Выборки без возвращения: дискретные аналоги некоторых непрерывных распределений.* — Зап. научн. семин. ПОМИ **311** (2004), 40–50.
2. A. Berred, V. Nevzorov, *Characterizations based on order statistics under sampling without replacement.* — J. Statist. Planning Infer. **139** (2009), 547–557.

Berred A., Nevzorov V. B. Samples without replacement: one property of distribution of range.

One characterization of distributions based on regressional property of range of sample is obtained for random sampling without replacement.

Faculté des Sciences et Techniques
B.P. 540, Le Havre Université
25, rue Lebon 76058
Le Havre Cedex, France
E-mail: alexandre.berred@univ-lehavre.fr

Поступило 2 ноября 2009 г.

С.-Петербургский государственный
университет, Университетеский пр. 28, Старый Петергоф,
198504 Санкт-Петербург, Россия
E-mail: vanev@mail.ru