В. Б. Невзоров

ОБ ОДНОМ ПРЕДЕЛЬНОМ СООТНОШЕНИИ МЕЖДУ ПОРЯДКОВЫМИ СТАТИСТИКАМИ И РЕКОРДАМИ

1. Введение

Пусть $X_1, X_2, \ldots,$ – независимые случайные величины с общей непрерывной функцией распределения F и $X_{1,n} \leqslant X_{2,n} \leqslant \ldots \leqslant X_{n,n}$, $n = 1, 2, \ldots,$ – соответствующие порядковые статистики. С порядковыми статистиками тесно связаны рекордные величины $L(n)$ и рекордные величины $X(n)$, которые определяются соотношениями:

$$L(1) = 1, \quad L(n+1) = \min\{j : X_j > X_{L(n)}\}, \quad n = 1, 2, \ldots \tag{1}$$

и

$$X(n) = X_{L(n)}, \quad n = 1, 2, \ldots. \tag{2}$$

Если быть точными, равенствами (1) и (2) определены верхние рекордные моменты и верхние рекордные величины. Если в (1) знак $>$ заменить на $<$, то будем иметь дело с нижними рекордами. Нетрудно заметить, что теория нижних рекордных моментов и величин совпадает с теорией верхних рекордов, поскольку для перехода от одних рекордов к другим достаточно вместо исходных случайных величин $X_1, X_2, \ldots$ рассмотреть последовательность $Y_1 = -X_1$, $Y_2 = -X_2, \ldots$. В дальнейшем мы будем говорить о рекордных моментах и рекордных величинах, имея в виду верхние рекорды.

Отметим, что рекордные величины соотношением

$$X(n) = X_{L(n),L(n)}, \tag{3}$$

связаны с максимальными порядковыми статистиками.

Работа поддержана Российским фондом фундаментальных исследований (грант 96-01-00549), грантом ведущих научных школ No. 96-15-96199 и Федеральной целевой программой интеграциия No. 326.53.

Подробно с теорией порядковых статистик, экстремальных порядковых статистик и рекордов можно познакомиться соответственно в книгах [1, 2] и обзоре [3].

Соотношение (3) связывает рекордные величины с максимальными членами вариационного ряда, построенного по выборке случайного объема $L(n)$. Ниже мы покажем, каким образом при исследовании рекордов можно использовать асиптотическую теорию экстремальных, но на этот раз – минимальных, порядковых статистик.

2. Совпадающие и различающиеся свойства рекордов и порядковых статистик

Мы уже привели соотношение (3), из которого можно заключить, что нас не должны удивлять совпадения ряда свойств рекордов и порядковых статистик. Приведем некоторые примеры.

Пример 1. Пусть $F(x) = \max\{0, 1 - \exp(-(x-a)/b)\}$. В [6] и [8] показано, что для экспоненциального распределения независимыми являются случайные величины

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

и соответственно

$$X(1), X(2) - X(1), \ldots .$$

Более того, независимость элементов как первой, так и второй последовательности при некоторых дополнительных ограничениях на функцию распределения F характеризуют именно экспоненциальное распределение.

Пример 2. Пусть $F(x) = \max\{0, 1 - \exp(-x/b)\}$. В этом случае совпадают по распределению нормированные спейсинги порядковых статистик $W_r = (n-r)(X_{r+1,n} - X_{r,n})$, $r = 0, 1, \ldots, n-1$, где $X_{0,n} = 0$, и общее распределение имеют межрекордные величины $X(1)$, $X(2) - X(1)$, $X(3) - X(2), \ldots$. На основе данного свойства рекордов и порядковых статистик также были получены различные характеризации экспоненциального распределения.

Эти свойства могли привести к предположению, что свойства порядковых статистик наследуются рекордами. Однако, уже следующий пример разрушает эту гипотезу.

Пример 3. Автор [5] показал, что если $0 < D(X(r)) < \infty$ для $r = i, j$, где $1 \leqslant i < j$, то для коэффициента корреляции $\rho(X(i), X(j))$ между рекордными величинами $X(i)$ и $X(j)$ справедливо неравенство

$$\rho(X(i), X(j)) \leqslant (i/j)^{1/2}, \tag{4}$$

причем равенство в (4) достигается тогда и только тогда, когда

$$F(x) = \max\{0, 1 - \exp(-(x-a)/b)\}, \quad x > a.$$

Казалось бы, аналогичный результат должен быть справедлив и для порядковых статистик, но в работе [7] было доказано, что если $0 < D(X_{r,n}) < \infty$ при некоторых $r = i, j$, где $1 \leqslant i < j \leqslant n$, то

$$\rho(X_{i,n}, X_{j,n}) \leqslant (i(n+1-j)/j(n-i+1))^{1/2}, \tag{5}$$

причем равенство (5) достигается тогда и только тогда, когда X_1 имеет равномерное распределение. Таким образом, характеризации *экспоненциального* распределения свойством максимального значения коэффициента корреляции между рекордами соответствует на этот раз характеризации *равномерного* распределения, основанная на аналогичном свойстве порядковых систем.

В четвертой части работы мы доказываем теорему, которая объясняет, почему возможны различные пары (экспоненциальное, экспоненциальное) и (равномерное, экспоненциальное) распределений, характеризуемых соответственно одним и тем же свойством порядковых статистик и рекордных величин. Предварительно вспомним некоторые результаты асимптотической теории экстремумов.

3. Предельные распределения минимальных порядковых статистик

Как известно (см., например [1]), имеется три типа предельных (при стремлении к бесконечности объема выборки n) распределений для надлежащим образом центрированных и нормированных минимумов $X_{1,n}$. Эти асимптотические функции распределения имеют вид

$$L_1(x) = L_{1,\gamma}(x) = 1 - \exp(-(-x)^{-\gamma}), \quad \text{если} \quad x < 0, \quad \text{и} \quad L_1(x) = 1, \quad \text{если} \quad x > 0, \quad \gamma > 0,$$
$$L_2(x) = L_{2,\gamma}(x) = 1 - \exp(-(-x)^{-\gamma}), \quad \text{если} \quad x > 0, \quad \text{и} \quad L_2(x) = 0, \quad \text{если} \quad x < 0, \quad \gamma > 0,$$
$$L_3(x) = 1 - \exp(-\exp x).$$

Мы говорим, что функция распределения F принадлежит области притяжения невырожденного предельного распределения минимум L и записываем этот факт, используя обозначение

$$F \in D(L),$$

если существуют такие последовательности $\{a_n\}$ и $\{b_n > 0\}$, что

$$1 - (1 - F(a_n + b_n x))^n \to L(x), \quad n \to \infty.$$

Отметим, что обе функции распределения, равномерная и экспоненциальная, встречающиеся в примерах 1–3, т.е.

$$F(x) - (x - a)/b, \quad a < x < b + a,$$

и

$$F(x) = 1 - \exp\{-(x - a)/b\}, \quad x > a,$$

принадлежат области притяжения предельной функции распределения $L_{2,1}(x) = 1 - \exp(-x)$. Оказывается, именно этот факт и помогает объяснить кажущееся противоречие с появлением различных распределений в примерах 1–3.

4. Основной результат

Нам потребуются следующие обозначения. Через

$$\{X(j),\ 1 \leqslant j \leqslant r;\ F\} \quad \text{и} \quad X_{j,n},\ 1 \leqslant j \leqslant r;\ F\},$$

будем обозначать рекордные величины $X(1),\ X(2), \ldots, X(r)$ и $X_{1,m}, X_{2,m}, \ldots, X_{r,n}$, соответствующие функции распределения F. Если

$$F(x) = \max\{0, 1 - \exp(-x)),$$

то экспоненциальные порядковые статистики и экспоненциальные рекордные величины обозначим $Z_{1,n},\ Z_{2,n}, \ldots, Z_{n,n}$ и $Z(1)$,

$Z(2),\ldots$ соответственно. В качестве вспомогательных независимых случайных величин, имеющих стандартное экспоненциальное распределение, будем использовать величины $\xi_1, \xi_2, \ldots$. Исходной функции распределения F сопоставим обратную функция F^1 и функцию $H(x) = F^1(1 - \exp(-x))$. Равенство

$$X \stackrel{d}{=} Y$$

означает, что случайные величины (или векторы) X и Y имеют одинаковое распределение.

Нам потребуются следующие представления для порядковых статистик и рекордов. Первые два из них формулируются для экспоненциальных порядковых статистик и экспоненциальных рекордов.

Представление 1.

$$\{Z_{r,n},\ 1 \leqslant r \leqslant n\} \stackrel{d}{=} \Big\{\sum_{j=1}^{r} \frac{\xi_j}{n-j+1},\ 1 \leqslant r \leqslant n\Big\}.$$

Представление 2.

$$\{Z(r),\ 1 \leqslant r \leqslant n\} \stackrel{d}{=} \Big\{\sum_{j=1}^{r} \xi_j,\ 1 \leqslant r \leqslant n\Big\}.$$

Следующие два представления справедливы для любой непрерывной функции распределения F.

Представление 3.

$$\{X_{r,n},\ 1 \leqslant r \leqslant n;\ F\} \stackrel{d}{=} \Big\{H\Big(\sum_{j=1}^{r} \frac{\xi_j}{n-j+1}\Big),\ 1 \leqslant r \leqslant n\Big\}.$$

Представление 4.

$$\{X(r),\ 1 \leqslant r \leqslant n;\ F\} \stackrel{d}{=} \Big\{H\Big(\sum_{j=1}^{r} \xi_j\Big),\ 1 \leqslant r \leqslant n\Big\}.$$

Основным результатом данной работы является следующая теорема.

Теорема. *Пусть F – непрерывная функция распределения и $F \in D(L)$, с центрирующими и нормирующими константами $\{a_n\}$ и $\{b_n\}$, где L – одна из функций распределений $L_{1,\gamma}(x)$, $L_{2,\gamma}(x)$ и $L_{3,\gamma}(x)$. Тогда для любого фиксированного $j = 1, 2, \ldots$ выполянется предельное соотношение*

$$\big\{X(j),\ 1 \leqslant j \leqslant r;\ L\big\}$$
$$\stackrel{d}{=} \lim \big\{(X_{j,n} - a_n)/b_n,\ 1 \leqslant j \leqslant r;\ F\big\}, \quad n \to \infty. \tag{6}$$

Доказательство теоремы. Если воспользоваться представлением 4, то совместная функция распределения рекордных величин $X(1),\ X(2), \ldots, X(r)$, соответствующих функций распределения L, может быть выражена с помощью стандартных экспоненциальных случайных величин $\xi_1,\, \xi_2, \ldots$ в следующем виде:

$$T(x_1, x_2, \ldots, x_r) = P\big\{X(1) < x_1,\ X(2) < x_2, \ldots, X(r) < x_r\big\}$$
$$= P\big\{\xi_1 < R(x_1),\ \xi_1 + \xi_2 < R(x_2), \ldots, \xi_1 + \xi_2 + \ldots + \xi_r < R(x_r)\big\}, \tag{7}$$

где
$$R(x) = -\log(1 - L(x)).$$

Совместное же распределение порядковых статистик $X_{1,n}$, $X_{2,n}, \ldots, X_{r,n}$ в аналогичной форме получим из представления 3:

$$T_n(x_1, x_2, \ldots, x_r)$$
$$= P\big\{X_{1,n} - a_n < x_1 b_n, X_{2,n} - a_n < x_2 b_n, \ldots, X_{r,n} - a_n < x_r b_n\big\}$$
$$= P\Big\{\xi_1 < R_n(x_1), \xi_1 + \xi_2 \frac{n}{n-1} < R_n(x_2), \ldots, \xi_1$$
$$+ \xi_2 \frac{n}{n-1} + \ldots + \xi_r \frac{n}{n-r+1} < R_n(x_r)\Big\}, \tag{8}$$

где
$$R_n(x) = -n \log\big(1 - F(a_n + b_n x)\big).$$

Сравнивая правые части (7) и (8), приходим к неравенству

$$|\, T(x_1, x_2, \ldots, x_r) - T_n(x_1, x_2, \ldots, x_r)\, | \leqslant S_1 + S_2,$$

где

$$S_1 = \sup_{x_1, \ldots, x_r} \Big| P\Big\{\xi_1 < R_n(x_1), \xi_1 + \xi_2 \frac{n}{n-1} < R_n(x_2), \ldots,$$

$$\ldots \xi_1 + \xi_2 \frac{n}{n-1} + \ldots + \xi_r \frac{n}{n-1+1} < R_n(x_r) \Big\}$$

$$-P\big\{\xi_1 < R_n(x_1), \xi_1 + \xi_2 < R_n(x_2), \ldots, \xi_1 + \xi_2 + \ldots + \xi_r < R_n(x_r)\big\}\Big|$$

и

$$S_2 = \sup_{x_1, \ldots, x_r} \mid P\{\xi_1 < R_n(x_1), \xi_1 + \xi_2 < R_n(x_2), \ldots, \ldots \xi_1$$

$$\ldots \xi_1 + \xi_2 + \ldots + \xi_r < R_n(x_r)\} - P\{\xi_1 < R_n(x_1),$$

$$\xi_1 + \xi_2 < R(x_2), \ldots, \xi_1 + \xi_2 + \ldots + \xi_r < R(x_r)\} \mid .$$

Очевидно, что для любого фиксированного $r = 1, 2, \ldots$ вектор

$$\Big(\xi_1, \xi_1 + \xi_2 \frac{n}{n-1}, \ldots, \xi_1 + \xi_2 \frac{n}{n-1} + \ldots + \xi_r \frac{n}{n-r+1}\Big)$$

сходится по распределению при $n \to \infty$ к вектору

$$(\xi_1, \xi_1 + \xi_2, \ldots, \xi_1 + \xi_2 + \ldots + \xi_r),$$

элементами которого являются последовательные суммы независимых случайных величин, имеющих стандартное экспоненциальное распределение. Следовательно, $S_1 \to 0$, если $n \to \infty$.

Вспомним теперь, что F принадлежит области притяжения предельной функции распределения минимумов L, причем соответствующими центрирующими и нормирующими константами являются последовательности $\{a_n\}$ и $\{b_n\}$. Это означает, что

$$1 - (1 - F(a_n + b_n x))^n \to L(x), \quad n \to \infty. \tag{9}$$

Из соотношения (9) следует, что

$$R_n(x) = -\log\big((1 - F(a_n + b_n x))^n\big) \to \log(1 - L(x)) = R(x), \quad n \to \infty,$$

и тогда $S_2 \to 0$, если $n \to \infty$. Утверждение теоремы доказано.

Замечание. Доказанная теорема объясняет, почему в примерах 1–3 экспоненциальные рекорды наследуют свойства не только экспоненциальных, но и равномерных порядковых статистик. Дело в том, что экспоненциальные и равномерные распределения, наряду со многими другими, принадлежат области притяжения экспоненциальной предельной функции распределения минимумов

$$L_{2,1}(x) = \max\{0, 1 - \exp(-x)\}.$$

Поэтому для широкого класса функционалов Φ из (6) следует сходимость

$$\Phi(\vec{X}_n) \to \Phi(\vec{X}), \quad n \to \infty, \tag{10}$$

где

$$\vec{X}_n = \{(X_{j,n} - a_n)/b_n,\ 1 \leqslant j \leqslant r;\ F\}, \quad \vec{X} = \{X(j),\ 1 \leqslant j \leqslant r;\ L\},$$

F – экспоненциальная или равномерная функция распределения, а L – экспоненциальная. Соотношение (10) очевидным образом справедливо для непрерывных ограниченных функционалов Φ, а для других функционалов оно требует дополнительной проверки и, возможно, некоторых дополнительных ограничений на исходную функцию распределения F.

Чтобы не создалось впечатление, что все возможные варианты применения полученного предельного соотношения (6) сводятся к парам распределений (равномерное, экспоненциальное) и (экспоненциальное, экспоненциальное), приведем еще один пример.

Пример. Пусть $F(x) = \exp(x)$, $x < 0$. Нетрудно убедиться, что $F \in D(L_3)$. В этом случае компоненты вектора

$$\big(X_{2,n} - X_{1,n},\, 2(X_{3,n} - X_{2,n}), \ldots , r(X_{r+1,n} - X_{r,n}); F\big)$$

являются независимыми и имеют стандартное экспоненциальное распределение для любых $r = 2, \ldots,$ и $n > r$. Из утверждения теоремы сразу получаем известный факт (см. [4]), состоящий в том, что для исходного распределения с функцией распределения $L_3(x) = 1 - \exp(-\exp(x))$ случайные величины

$$X(2) - X(1),\, 2(X(3) - X(2)), \ldots , r(X(r+1) - X(r)), \quad r = 2, 3, \ldots ,$$

также независимы и имеют стандартное экспоненциальное распределение.

Литература

1. Я. Галамбош, *Асимптотическая теория экстремальных порядковых статистик.* Наука, М. (1984).
2. Г. Дэйвид, *Порядковые статистики.* Наука, М. (1979).
3. В. Б. Невзоров, *Рекорды.* — Теор. вероятн. и ее примен. **32, No. 2** (1987), 219–252.
4. R. L. Houchens, *Record value theory and inference.* Ph. D. Thesis, University of California, Riverside (1984).

5. V. Nevzorov, *A characterization of exponential distributions by correlation between records.* — Mathematical Methods of Statistics **1** (1992), 49–54.
6. P. V. Sukhatme, *Tests of significance for samples of the* χ^2 *population with two degrees of freedom.* — Ann. Eugenics **8** (1937), 52–56.
7. G. J. Szekely, T. F. Mori, *An extremal property of rectangular distributions.* — Stat. Probab. Lett. **3** (1975), 107–109.
8. M. N. Tata, *On outstanding values in a sequence of random variables.* — Z. Wahrscheinlichkeitsth. verw. Geb. **12** (1969), 9–20.

Nevzorov V. B. On a limit relation between order statistics and records.

Let a distribution function (d.f.) F belong to the domain of minimal attraction of some d.f. L. It is shown that a limit distribution of suitably normalized order statistics $X_{k,n}$ corresponding to d.f. F coincides with the distribution of the record value $X(k)$ corresponding to d.f. L.

С.-Петербургский
государственный университет

Поступило 11 декабря 1997 г.