УДК 519.2 Вестник СПбГУ. Сер. 1. Т. 2 (60). 2015. Вып. 1

АНАЛОГ РЕКОРДНОЙ F^α-СХЕМЫ ДЛЯ ДИСКРЕТНЫХ РАСПРЕДЕЛЕНИЙ*

В. Б. Невзоров

Санкт-Петербургский государственный университет,
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7-9

Имеется последовательность независимых случайных величин $X_1, X_2, \ldots$, принимающих целые неотрицательные значения. Рассматриваются случайные индикаторы $\eta_1, \eta_2, \ldots$, определенные таким образом, что $\eta_n = 1$, если значение n $(n = 0, 1, \ldots)$ фиксируется как рекордное в последовательности X-ов, и $\eta_n = 0$, иначе. Известно, что если исходные величины одинаково распределены, то индикаторы независимы. Исследуются последовательности неодинаково распределенных X-ов, для которых эти рекордные индикаторы остаются независимыми. Библиогр. 6 назв.

Ключевые слова: дискретные распределения, рекордные значения, рекордные индикаторы, рекордная F^α-схема.

1. Введение. В классической теории рекордных величин рассматриваются последовательности $X_1, X_2, \ldots$ независимых случайных величин (с.в.), имеющих одинаковую функцию распределения $F(x)$. Рекордные моменты $L(1), L(2), \ldots$ и рекордные величины $X(1), X(2), \ldots$ определяются следующим образом:

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

и

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

Соотношения (1) и (2) задают верхние рекордные моменты и верхние рекордные величины. Аналогично (с заменой в (1) неравенства $X_j > X_{L(n)}$ на неравенство $X_j < X_{L(n)}$) определяются нижние рекорды. Для дискретных распределений также вводят слабые рекордные величины. В этом случае повторение предыдущего рекордного значения засчитывается как новый рекорд. В теории рекордов обычно по отдельности рассматривают два типа распределений — непрерывных и дискретных. При исследовании этих моделей используются разные методы. Подробно с этими методами можно познакомиться в монографиях [1–3].

Мы рассмотрим два разных типа рекордных индикаторов, которые существенным образом используются при изучении рекордов. Пусть $X_1, X_2, \ldots$ — независимые с.в. с общей непрерывной функцией распределения $F(x)$. Введем случайные индикаторы $\xi_2, \xi_3, \ldots$ следующим образом:

$$\xi_n = \begin{cases} 1, & \text{если } X_n \text{ является верхней рекордной величиной,} \\ & \text{т. е. если } X_n > \max(X_1, X_2, \ldots, X_{n-1}), \\ 0, & \text{иначе}. \end{cases}$$

Определим также индикатор $\xi_1 = 1$, поскольку с.в. X_1 является рекордной.

*Работа выполнена при финансовой поддержке СПбГУ (проект 6.38.672.2013) и РФФИ (грант № 13-02-00338).

Реньи [4] в 1962 г. показал, что в этой ситуации (когда функция распределения F непрерывна) индикаторы $\xi_1, \xi_2, \ldots$ независимы и

$$P\{\xi_n = 1\} = 1/n, \quad n = 1, 2, \ldots \tag{3}$$

Этот результат позволяет представить число рекордов $N(n)$ в наборе $X_1, X_2, \ldots, X_n$ в виде суммы независимых слагаемых

$$N(n) = \xi_1 + \xi_2 + \ldots + \xi_n,$$

что, например, практически сразу позволяет показать асимптотическую нормальность при $n \to \infty$ с.в.:

$$(N(n) - \ln n)/(\ln n)^{1/2}.$$

Далее, учитывая справедливость соотношения

$$P\{L(m) > n\} = P\{N(n) < m\}, \tag{4}$$

можно уже доказать, что к нормальному предельному закону сходится по распределению и последовательность с.в.

$$(\ln L(n) - n)/n^{1/2}.$$

В связи с этой важной ролью индикаторов $\xi_1, \xi_2, \ldots$ при исследовании распределений (включая и предельные) числа рекордов $N(n)$ и рекордных моментов $L(n)$ возникает естественное желание попытаться расширить класс исходных случайных величин, для которых соответствующие индикаторы оставались бы независимыми. Здесь сразу следует отметить так называемую F^{α}-схему, частный случай которой был введен в рассмотрение в работе Йенга [5], и в дальнейшем этот случай был существенно обобщен (см., например, [3, лекция 25]).

В этой схеме исходные независимые с.в. $X_1, X_2, \ldots$ имеют функции распределения

$$F_k(x) = P\{X_k < x\} = F^{\alpha(k)}(x),$$

где $F(x)$ — некоторая исходная непрерывная функция распределения, а $\alpha(1), \alpha(2), \ldots$ — произвольные положительные числа. Было показано, что в рамках данной модели индикаторы рекордов $\xi_1, \xi_2, \ldots$ остаются независимыми и

$$P\{\xi_n = 1\} = \frac{\alpha(n)}{\alpha(1) + \cdots + \alpha(n)}, \quad n = 1, 2, \ldots$$

В ситуации же, когда изучаются рекорды в последовательностях дискретных с.в., важную роль играет другой тип рекордных индикаторов. Рассмотрим для простоты обозначений случай, когда исходные случайные величины независимы, одинаково распределены и принимают целые неотрицательные значения с вероятностями $p_k = P\{X_j = k\}$, $k = 0, 1, \ldots$, $j = 1, 2, \ldots$. Введем с.в. $\eta_0, \eta_1, \ldots$, определенные таким образом, что $\eta_n = 1$, если значение n фиксируется как рекордное в последовательности наблюдаемых значений с.в. $X_1, X_2, \ldots$, и $\eta_n = 0$, иначе. Шоррок [6] показал, что в этом случае таким образом введенные индикаторы также независимы и

$$P\{\eta_n = 1\} = 1 - P\{\eta_n = 0\} = P\{X = n\}/P\{X \geq n\}, \quad n = 0, 1, 2, \ldots \tag{5}$$

Это свойство индикаторов η_n позволило существенно упростить исследование дискретных рекордных величин $X(n)$. Действительно, в этой ситуации можно воспользоваться соотношением

$$P\{X(n) > m\} = P\{S(m) < n\}, \tag{6}$$

где $S(m) = \eta_0 + \eta_1 + \ldots + \eta_m$ — число рекордов, не превосходящих значения m.

Например, в случае геометрического распределения с вероятностями

$$P\{X_k = m\} = (1-p)\, p^m, \quad k = 1, 2, \ldots\,, \quad m = 0, 1, \ldots\,, \quad 0 < p < 1,$$

можно сразу получить, что

$$P\{\eta_n = 1\} = 1 - P\{\eta_n = 0\} = 1 - p, \quad n = 0, 1, 2, \ldots\,.$$

Отсюда и из соотношения (6) сразу следует, что случайные величины

$$\frac{S(n) - n(1-p)}{\sqrt{np\,(1-p)}} \quad \text{и} \quad \frac{(1-p)X(n) - n}{\sqrt{np}}$$

имеют в пределе при $n \to \infty$ стандартное нормальное распределение.

Возникает вопрос: можно ли в случае дискретных распределений существенно расширить множество последовательностей X-ов, для которых индикаторы $\eta_0, \eta_1, \ldots$ оставались бы независимыми? Эту проблему будем решать во второй части статьи.

2. Основные результаты. Рассмотрим вначале простейшую ситуацию, когда имеем дело с набором независимых неотрицательных целочисленных случайных величин $Y, X_2, X_3, \ldots$, в котором с.в. $X_2, X_3, \ldots$ имеют одинаковое распределение с вероятностями

$$f_n = P\{X_k = n\}, \quad n = 0, 1, 2, \ldots\,, \quad k = 2, 3, \ldots\,,$$

а распределение Y задается набором вероятностей

$$g_n = P\{Y = n\}, \quad n = 0, 1, 2, \ldots\,.$$

Обозначим также для значений $n = 1, 2, \ldots$ вероятности $F(n) = P\{X_k < n\}$, $k = 2, 3, \ldots$, и $G(n) = P\{Y < n\}$. Найдем условия, при которых в этом случае независимыми будут рекордные индикаторы $\eta_0, \eta_1, \eta_2, \ldots$.

Предполагая, что эти индикаторы независимы, рассмотрим при $k = 1, 2, \ldots$ соотношения

$$P\{\eta_0{=}0, \eta_1{=}0, \ldots, \eta_{k-1}{=}0, \eta_k{=}1\} = P\{\eta_0{=}0, \eta_1{=}0, \ldots, \eta_{k-1}{=}0\}\, P\{\eta_k{=}1\}\,. \tag{7}$$

Справедливы равенства

$$P\{\eta_0 = 0, \eta_1 = 0, \ldots, \eta_{k-1} = 0\} = P\{Y \geq k\} = 1 - G(k), \tag{8}$$

$$P\{\eta_0 = 0, \eta_1 = 0, \ldots, \eta_{k-1} = 0, \eta_k = 1\} = P\{Y = k\} = g_k, \tag{9}$$

$$\begin{aligned} P\{\eta_k = 1\} = P\{Y = k\} + P\{Y < k\}P\{X_1 = k\} + \\ + P\{Y < k\}P\{X_1 < k\}P\{X_2 = k\} + \ldots = g_k + \frac{f_k\, G(k)}{1 - F(k)}. \end{aligned} \tag{10}$$

Из соотношений (7)–(10) получаем, что

$$g_k = (1 - G(k)) \left(g_k + \frac{f_k\, G(k)}{1 - F(k)} \right),$$

и следовательно,

$$\frac{g_k}{1 - G(k)} = \frac{f_k}{1 - F(k)}, \quad k = 1, 2, \dots\ , \tag{11}$$

т. е. две дискретные функции интенсивности отказов совпадают при $k = 1, 2, \dots$

Равенства (11) можно переписать в виде

$$\frac{1 - G(k+1)}{1 - G(k)} = \frac{1 - F(k+1)}{1 - F(k)}, \quad k = 1, 2, \dots\ . \tag{12}$$

Перемножая, соответственно, выражения (при $k = 1, 2, \dots, n$), входящие в левые и правые части равенств (12), получаем, что

$$1 - F(n) = \lambda\ (1 - G(n))\,, \quad n = 1, 2, \dots, \tag{13}$$

где $\lambda = (1 - f_0)/(1 - g_0)$.

Можно отметить также, что из (13) следуют равенства

$$\frac{f_k}{1 - F(k)} = \frac{g_k}{1 - G(k)}, \quad k = 1, 2, \dots\ .$$

Рассмотрев приведенный частный случай, можно предположить, что если перейти к набору независимых случайных величин, принимающих значения $0, 1, 2, \dots$ с вероятностями

$$p_n(k) = P\{X_n = k\}, \quad k = 0, 1, 2, \dots\ , \quad n = 1, 2, \dots\ ,$$

имеющих соответствующие функции распределения $F_n(k) = p_n(0) + p_n(1) + \ldots + p_n(k)$ и дискретные интенсивности отказа $R_n(k) = p_n(k)/(1 - F_n(k))$, которые совпадают при всех $k = 1, 2, \dots$, то получим независимые рекордные индикаторы $\eta_0, \eta_1, \eta_2, \dots$. Проверим справедливость этого предположения.

В данной ситуации соотношения (8)–(10) перепишутся в следующем виде:

$$P\{\eta_0 = 0, \eta_1 = 0, \dots, \eta_{k-1} = 0\} = P\{X_1 \geq k\} = 1 - F_1(k), \tag{14}$$

$$P\{\eta_0 = 0, \eta_1 = 0, \dots, \eta_{k-1} = 0, \eta_k = 1\} = P\{X_1 = k\} = p_1(k), \tag{15}$$

$$P\{\eta_k = 1\} = p_1(k) + F_1(k)p_2(k) + \ldots + F_1(k)F_2(k) \cdots F_{n-1}(k)p_n(k) + \dots\ . \tag{16}$$

Для проверки независимости индикаторов $\eta_0, \eta_1, \eta_2, \dots$, учитывая, что имеем дело с двухточечными случайными величинами, достаточно убедиться, что при любом $k = 1, 2, 3, \dots$ выполняются равенства

$$p_1(k) = (1 - F_1(k)) \Big(p_1(k) + F_1(k)p_2(k) + \ldots + F_1(k)F_2(k) \cdots F_{n-1}(k)p_n(k) + \dots \Big). \tag{17}$$

Нетрудно проверить, что для одинаково распределенных X-ов соотношение (17) выполняется. Обозначим

$$T_n(k) = p_n(k) + F_n(k) \Big(p_{n+1}(k) + F_{n+1}(k)p_{n+2}(k) + F_{n+1}(k)F_{n+2}(k)p_{n+3}(k) + \dots \Big).$$

Соотношение (17) перепишем в виде

$$p_1(k) = (1 - F_1(k))\Big(p_1(k) + F_1(k)T_2(k)\Big), \tag{18}$$

откуда следует, что

$$p_1(k)/(1 - F_1(k)) = T_2(k). \tag{19}$$

Напомним, что по условию

$$\frac{p_1(k)}{1 - F_1(k)} = \frac{p_2(k)}{1 - F_2(k)}, \quad k = 1, 2, \dots\ . \tag{20}$$

Поэтому из равенств (19) и (20) получаем, что

$$p_2(k) = (1 - F_2(k))\, T_2(k), \quad k = 1, 2, \dots\ . \tag{21}$$

Равенство (21) можно теперь переписать в виде

$$p_2(k) = (1 - F_2(k))\Big(p_2(k) + F_2(k)T_3(k)\Big). \tag{22}$$

Сравнивая соотношения (18) и (22), видим, что удалось избавиться от зависимости исходного равенства (17) от характеристик с.в. X_1. Продолжая дальше данную процедуру и беря в качестве начального (вместо равенства (18)) соотношение (22), показываем, что таким же образом и исходное распределение с.в. X_2 приводит нас уже к проверке очередного соотношения

$$p_3(k) = (1 - F_3(k))\Big(p_3(k) + F_3(k)T_4(k)\Big). \tag{23}$$

Используя и далее предложенный метод, мы можем доказать, что если зафиксировать произвольное сколь угодно большое значение n и рассмотреть последовательность целочисленных независимых случайных величин $X_1, X_2, \dots, X_n, X_{n+1}, \dots$, среди которых с.в. $X_n, X_{n+1}, \dots$ одинаково распределены, а величины $X_1, X_2, \dots, X_n$ имеют интенсивности отказа $R_m(k)$, $m = 1, \dots, n$, которые совпадают при $k = 1, 2, \dots$, то рекордные индикаторы $\eta_0, \eta_1, \eta_2, \dots, \eta_n, \eta_{n+1}, \dots$ независимы.

Существенно более громоздкие выкладки, связанные с прямой подстановкой в (17) вероятностей $p_m(k)$ и $F_m(k)$ случайных величин, у которых интенсивности отказа $R_m(k) = p_m(k)/(1 - F_m(k))$, $m = 1, 2, \dots$, совпадают при всех $k = 1, 2, \dots$, позволяют показать, что рекордные индикаторы и в этом случае независимы.

Таким образом, в качестве «дискретного» аналога F^α-схемы можно предложить следующую модель. Берутся независимые с.в. $X_1, X_2, \dots$, принимающие значения $0, 1, \dots$ с вероятностями $p_k(0), p_k(1), \dots$, $k = 1, 2, \dots$, для которых при любых значениях индексов k и r и для всех $m = 1, 2, \dots$ выполняются равенства

$$\frac{p_k(m)}{1 - F_k(m)} = \frac{p_r(m)}{1 - F_r(m)}.$$

В качестве примера одной из таких последовательностей можно привести набор геометрически распределенных с.в. $X_1, X_2, \dots$ с вероятностями

$$\begin{aligned} P\{X_n = 0\} &= 1 - p^n, \\ P\{X_n = m\} &= (1 - p)\, p^{n+m-1}, \quad m = 1, 2, \dots\ , \quad 0 < p < 1. \end{aligned}$$

Видим, что в этом случае для каждого $n = 1, 2, \ldots$ и любого $m = 1, 2, \ldots$ выполняется равенство

$$P\{X_n = m\}/P\{X_n > m\} = 1 - p$$

и данная последовательность представляет собой дискретный аналог F^α-схемы.

Литература

1. *Ahsanullah M., Nevzorov V. B.* Ordered random variables. New York: Nova Science Publishers, 2001. 412 p.

2. *Arnold B. C., Balakrishnan N., Nagaraja H. N.* Records. New York: John Wiley and Sons, 1998. 312 p.

3. *Невзоров В. Б.* Рекорды. Математическая теория. М.: ФАЗИС, 2000. 244 с.

4. *Renyi A.* Theorie des elements saillants d'une suite d'observations // Colloquium on Combinatorial Methods in Probability Theory, Mathematisk Institut, Aarhus University. Aarhus, Denmark, 1962. P. 104–117.

5. *Yang M. C. K.* On the distribution of the inter-record times in an increasing population // Journal of Applied Probability, 1981. Vol. 12. P. 148–154.

6. *Shorrock R. W.* On record values and record times // Journal of Applied Probability, 1972. Vol. 9. P. 316–326.

Статья поступила в редакцию 23 октября 2014 г.

Сведения об авторе

Невзоров Валерий Борисович — доктор физико-математических наук, профессор; probabil@pisem.net

AN ANALOGUE OF THE F^α-SCHEME FOR DISCRETE DISTRIBUTIONS

Valery B. Nevzorov

St.Petersburg State University, Universitetskaya nab., 7-9, St.Petersburg, 199034, Russian Federation; probabil@pisem.net

Let $X_1, X_2, \ldots$ be a sequence of independent random variables with nonnegative integer values. Define random indicators $\eta_0, \eta_1, \ldots$ as follows: $\eta_n = 1$, if n $(n = 0, 1, \ldots)$ is a record value in the sequence of X's, and $\eta_n = 0$, otherwise. It is known that these indicators are independent if the initial X's are identically distributed. We consider sequences of non-identically distributed X's, for which record indicators $\eta_0, \eta_1, \ldots$ are independent. Refs 6.

Keywords: discrete distributions, record values, record indicators, record F^α-scheme.

References

1. Ahsanullah M., Nevzorov V. B., "Ordered random variables" (Nova Science Publishers, New York, 2001).

2. Arnold B. C., Balakrishnan N., Nagaraja H. N., "Records" (John Wiley and Sons, New York, 1998).

3. Nevzorov V. B., *Records: Mathematical Theory* (FAZIS, Moscow, 2000) [in Russian].

4. Renyi A., "Theorie des elements saillants d'une suite d'observations", *Colloquium on Combinatorial Methods in Probability Theory* 104–117 (Mathematisk Institut, Aarhus University. Aarhus, Denmark, 1962).

5. Yang M. C. K., "On the distribution of the inter-record times in an increasing population", *Journal of Applied Probability* **12**, 148–154 (1981).

6. Shorrock R. W., "On record values and record times", *Journal of Applied Probability* **9**, 316–326 (1972).