МАТЕМАТИКА

УДК 519.2
MSC 62G32

Об одной проблеме оптимального выбора рекордных величин

И. В. Бельков, В. Б. Невзоров

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

Для цитирования: *Бельков И. В., Невзоров В. Б.* Об одной проблеме оптимального выбора рекордных величин // Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия. 2018. Т. 5 (63). Вып. 2. С. 179–188. https://doi.org/10.21638/11701/spbu01.2018.201

В классической задаче оптимального выбора («задача о разборчивой невесте», «проблема секретаря») обсуждалась процедура, позволяющая с максимально возможной вероятностью, последовательно получая результаты наблюдений случайных величин и в нужный момент останавливаясь на одном из них, выбрать последнее рекордное значение в наборе из n случайных величин. В предлагаемой статье исследуются соответствующие процедуры выбора в последовательности из n случайных величин начальной точки отсчета рекордных значений, позволяющие максимизировать математическое ожидание суммарного числа верхних и нижних рекордов.

Ключевые слова: рекордные моменты, рекордные величины, среднее число рекордов, равномерное распределение, задача оптимального выбора.

Пусть $X_1, X_2, \ldots$ — последовательность независимых случайных величин (с. в.), имеющих непрерывную функцию распределения (ф. р.) $F(x)$. Определим для этой последовательности верхние рекордные моменты $L(n)$ и верхние рекордные величины $X(n), n = 1, 2, \ldots$ Зададим эти случайные величины следующими соотношениями:

$$\begin{gathered} L(1) = 1, \quad X(1) = X_1; \\ L(n) = \min\{j : X_j > X(n-1)\}, \quad X(n) = X_{L(n)}, \quad n = 2, 3, \ldots \end{gathered} \tag{1}$$

Аналогичным образом можно определить нижние рекордные моменты $l(n)$ и нижние рекордные величины $x(n), n = 1, 2, \ldots$:

$$l(1) = 1, \quad x(1) = X_1;$$
$$l(n) = \min\{j : X_j < x(n-1)\}, \quad x(n) = X_{l(n)}, \quad n = 2, 3, \ldots \tag{2}$$

Для любого $n = 1, 2, \ldots$ и набора $X_1, X_2, \ldots, X_n$ рассмотрим также случайные величины $M(n)$ и $m(n)$, представляющие собой, соответственно, число верхних и число нижних рекордов в наборе $X_1, X_2, \ldots, X_n$, и их суммы обозначим $N(n) = M(n)+m(n), n = 1, 2, \ldots$ Отметим также, что с. в. X_1 является одновременно первой верхней и первой нижней рекордной величиной.

Введем в рассмотрение также с. в. $\xi_1, \xi_2, \ldots$ и $\eta_1, \eta_2, \ldots$, представляющие собой индикаторы верхних и нижних рекордов. Для них справедливы равенства

$$P\{\xi_1 = 1\} = P\{\eta_1 = 1\} = 1,$$

$$P\{\xi_k = 1\} = P\{X_k > \max(X_1, \ldots, X_{k-1})\} = \frac{1}{k}, \quad k = 2, 3, \ldots,$$

$$P\{\eta_k = 1\} = P\{X_k < \min(X_1, \ldots, X_{k-1})\} = \frac{1}{k}, \quad k = 2, 3, \ldots$$

Отметим, что $M(n) = \xi_1 + \xi_2 + \ldots + \xi_n$ и $m(n) = \eta_1 + \eta_2 + \ldots + \eta_n$, $n = 1, 2, \ldots$

Справедливы следующие равенства для математических ожиданий $a_n = EM(n)$, $b_n = Em(n)$ и $c_n = EN(n)$:

$$a_n = b_n = 1 + \frac{1}{2} + \ldots + \frac{1}{n}, \quad c_n = a_n + b_n = 2\left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right), \quad n = 1, 2, \ldots, \tag{3}$$

из которых, в частности, следуют соотношения

$$a_n \sim \ln n, \quad b_n \sim \ln n, \quad c_n \sim 2\ln n, \quad n \to \infty. \tag{4}$$

Будем также использовать специальные обозначения $U_1, U_2, \ldots$ для случайных величин, имеющих $U([0,1])$-равномерное распределение на интервале $[0,1]$: $U_1 = U(1) < U(2) < \ldots$ для верхних рекордных величин и $U_1 = u(1) > u(2) > \ldots$ для нижних рекордов, построенных по этой последовательности.

Различные результаты для рекордных моментов и рекордных величин можно найти, например, в монографиях [1–4], в статьях [5–8].

С верхними рекордными величинами связана классическая задача оптимального выбора, также называемая «*задачей о разборчивой невесте*» или «*проблемой секретаря*», которую можно сформулировать следующим образом.

Имеется набор из n случайных величин $X_1, X_2, \ldots, X_n$ и соответствующий им набор рекордных величин $X(1) = X_1 < X(2) < \ldots < X(M) = \max\{X_1, X_2, \ldots, X_n\}$. Последовательно получаем наблюдаемые значения x_1, x_2, $\ldots$ исходных X_1, X_2, $\ldots$ Получив очередное рекордное наблюдение x_k, которое больше всех предыдущих значений $x_1, x_2, \ldots, x_{k-1}$, нужно решить, останавливаться ли на этом наблюдении, предполагая, что оно и является значением с. в. $X(M)$, или, расставшись с этим наблюдением (и уже не имея возможности в этом случае к нему вернуться), продолжить процесс выбора в надежде все же выйти позже на максимальное из всех n наблюдений. По сути дела, максимизируется вероятность p_n угадать момент появления последнего рекорда среди величин $X_1, X_2, \ldots, X_n$.

Оптимальная стратегия, позволяющая решить эту проблему (см., например, [9]), заключается в том, что нужно при фиксированном числе n элементов выборки получить, но проигнорировать какое-то число $r = r(n)$ наблюдений $x_1, x_2, \ldots, x_r$, а затем выбрать первое же значение x_s, $s > r$, которое превышает все предыдущие. Имеются таблицы, позволяющие для значений $n = 2, 3, \ldots$ находить величины $r = r(n)$ и вероятности p_n. Отметим лишь, что $r(n) \sim n/e$ и $p_n \to 1/e = 0{,}367879\ldots$ при $n \to \infty$.

В статье [10] была рассмотрена задача оптимального выбора, в которой интерес представлял вопрос о возможности увеличить математическое ожидание числа верхних рекордов среди с. в. $X_1, X_2, \ldots, X_n$, если применить процедуру, использованную в задаче о разборчивой невесте, т. е. если начинать отсчет рекордов не с первого элемента последовательности, а лишь проведя какое-то число наблюдений и дождавшись появления некоторого достаточно малого по величине наблюдаемого значения x_r.

Сначала рассматриваются исходные наборы из n независимых с. в. $U_1, U_2, \ldots, U_n$, имеющих $U([0,1])$-равномерное распределение. Для каждого фиксированного $n = 3, 4, \ldots$ дело сводится к нахождению единственного на интервале $(0,1)$ корня уравнения

$$\sum_{k=1}^{n-1} \frac{t^k(n)}{k} = 1, \quad n = 3, 4, \ldots \tag{5}$$

Оптимизационная процедура в этом случае состоит в следующем. Наблюдается значение x_1 с. в. U_1. Если $x_1 < t(n)$, то $U_1 = x_1$ выбирается в качестве первого рекордного значения, с которого и начинается отсчет рекордов. Если $x_1 \geqslant t(n)$, отбрасываем это значение и наблюдаем значение x_2 второй случайной величины U_2. Здесь уже требуется знать корень $t(n-1)$ уравнения

$$\sum_{k=1}^{n-2} \frac{(t(n-1))^k}{k} = 1.$$

Если $x_2 < t(n-1)$, начальной точкой отсчета рекордов будет это самое значение x_2. Если же $x_2 \geqslant t(n-1)$, переходим уже к с. в. U_3, не имея в дальнейшем возможности вернуться к величинам $U_1 = x_1$ и $U_2 = x_2$. Как только происходит первое из событий

$$\{U_1 < t(n)\}, \{U_1 \geqslant t(n), U_2 < t(n-1)\}, \ldots,$$
$$\{U_1 \geqslant t(n), U_2 \geqslant t(n-1), \ldots, U_{r-1} \geqslant t(n-r+2), U_r < t(n-r+1)\}, \ldots,$$

выбираем соответствующее значение x_r случайной величины U_r в качестве начальной точки отсчета рекордов. Если при этом дело доходит до появления события

$$\{U_1 \geqslant t(n), U_2 \geqslant t(n-1), \ldots, U_{n-3} \geqslant t(4), U_{n-2} \geqslant t(3)\},$$

уже приходится ограничиваться для подсчета рекордов двумя с. в. U_{n-1} и U_n. В работе [10] были найдены величины $t(n)$ для различных n и максимально возможные значения математических ожиданий числа верхних рекордов, которые можно получить в результате такой оптимизационной процедуры. Отметим, что при переходе от исходных равномерных с. в. $U_1, U_2, \ldots, U_n$ к величинам $X_1, X_2, \ldots, X_n$, имеющим произвольную непрерывную ф. р. $F(x)$, изменяются лишь критические значения $t(n)$, $n = 3, 4, \ldots$, вместо которых нужно будет брать величины $G(t(n))$, где $G(x) = F^{-1}(x)$ — функция, обратная функции распределения $F(x)$. Соответствующие максимальные значения математических ожиданий чисел верхних рекордов среди с. в. $X_1, X_2, \ldots, X_n$ остаются такими же, какими были получены для наборов $U_1, U_2, \ldots, U_n$.

Естественным образом описанная для числа верхних рекордов оптимизационная процедура переносится на аналогичную ситуацию с нижними рекордами. Действительно, достаточно от набора $X_1, X_2, \ldots, X_n$ перейти к с. в. $-X_1, -X_2, \ldots, -X_n$, чтобы можно было переформулировать уже имеющиеся результаты для верхних рекордов среди X_1, X_2, $\ldots$ на случай нижних рекордов. В этом случае нужно воспользоваться уже полученными в ситуации с верхними рекордами значениями $t(3) > t(4) > \ldots$ и, имея набор с. в. $X_1, X_2, \ldots, X_n$ с непрерывной ф. р. $F(x)$, ждать появления первого из событий

$$B_1 = \{X_1 \geqslant G(1 - t(n))\}, B_2 = \{X_1 < G(1 - t(n)), X_2 \geqslant G(1 - t(n-1))\}, \ldots,$$

$$B_r = \{X_1 < G(1 - t(n)), X_2 < G(1 - t(n-1)), \ldots, \\ X_{r-1} < G(1 - t(n-r+2)), X_r \geqslant G(1 - t(n-r+1))\}, \ldots$$

При появлении первого в этом ряду события B_m надо наблюдаемое значение x_m соответствующей этому событию с. в. X_m объявить первым рекордным и начать отсчет рекордов именно в этот момент. Оптимальным образом полученное в результате такой немного видоизмененной процедуры максимальное среднее значение числа нижних рекордов в наборе $X_1, X_2, \ldots, X_n$ не будет отличаться от максимального математического ожидания числа верхних рекордов.

В связи с этими двумя уже исследованными в работе [10] ситуациями, когда по отдельности рассматриваются верхние и нижние рекорды, интерес представляет описание соответствующей процедуры нахождения начальной точки отсчета рекордов в случае, когда нужно максимизировать математическое ожидание суммарного числа как верхних, так и нижних рекордных величин в наборе $X_1, X_2, \ldots, X_n$.

Отметим, что ситуации, когда на равных основаниях интерес представляют оба описанных выше типа рекордов, нередки в различных областях человеческой деятельности. Например, фиксируются как рекордные отрицательные, так и рекордные положительные температуры воздуха за все годы их измерений в той или иной стране. Привлекают внимание статистиков рекордно высокие и рекордно низкие урожаи той или иной сельскохозяйственной культуры, а также верхние и нижние рекордные значения уровней различных водоемов.

Вернемся к исходному набору независимых с. в. $X_1, X_2, \ldots, X_n$, имеющих общую непрерывную ф. р. $F(x)$. Математическое ожидание числа верхних и нижних рекордов среди этих величин задается равенством

$$c_n = EN(n) = 2\left(1 + \frac{1}{2} + \ldots + \frac{1}{n}\right), \quad n = 1, 2, \ldots$$

Какой должна быть процедура, чтобы, последовательно получая наблюдения $x_1, x_2, \ldots$, остановиться в нужный момент на наблюдаемом значении x_r некоторой соответствующим образом выбранной с. в. X_r и уже с этого места начать отсчет как верхних, так и нижних рекордных величин? Насколько получаемое таким образом среднее значение суммарного числа рекордов будет больше c_n?

Выше уже было показано, что достаточно нужную процедуру описать на примере последовательности независимых с. в. $U_1, U_2, \ldots, U_n$, имеющих $U([0,1])$-равномерное распределение.

Предположим, что нашли значение x_1 первой в этом наборе с. в. U_1. Нужно решить, взять ли это значение как первое рекордное или, отбросив это наблюдение и уже не имея возможности в дальнейшем к нему вернуться, перейти к с. в. U_2, найти ее значение x_2 и уже здесь решать, останавливаться ли на U_2 или, отбрасывая возможность к нему вернуться, двигаться дальше, получая значения очередных случайных величин из данного набора.

Итак, допустим, что получено значение x, $0 \leqslant x \leqslant 1$, с. в. U_1. Если это наблюдение принять за первый рекорд (верхний и нижний), математическое ожидание $N(x,n)$ суммарного числа рекордов в наборе $x, U_2, U_3, \ldots, U_n$ будет иметь вид

$$\begin{aligned} N(x,n) = 2 + \int_0^x du + \int_x^1 du + \int_0^x (1-u)du + \int_x^1 udu + \ldots + \\ + \int_0^x (1-u)^{n-2}du + \int_x^1 u^{n-2}du = \\ = 3 + 1 + \frac{2}{3} + \ldots + \frac{2}{n-1} - \left(\frac{x^2}{2} + \frac{x^3}{3} + \ldots + \frac{x^{n-1}}{n-1}\right) - \\ - \left(\frac{(1-x)^2}{2} + \frac{(1-x)^3}{3} + \ldots + \frac{(1-x)^{n-1}}{n-1}\right) = \\ = 1 + c_{n-1} - \left(\frac{x^2}{2} + \frac{x^3}{3} + \ldots + \frac{x^{n-1}}{n-1}\right) - \\ - \left(\frac{(1-x)^2}{2} + \frac{(1-x)^3}{3} + \ldots + \frac{(1-x)^{n-1}}{n-1}\right). \end{aligned} \quad (6)$$

Если решаем отбросить это наблюдение, должны перейти к набору $U_2, U_3, \ldots, U_n$, в котором среднее суммарное число наблюдений равно c_{n-1}. Таким образом, для принятия решения, взять ли x_1 в качестве первого рекордного значения или, отказавшись от него и не имея уже возможности в дальнейшем к нему вернуться, перейти к следующей с. в. U_2, нужно выяснить, какая из величин $N(x,n)$ и c_{n-1} больше. Дело сводится к нахождению на интервале $[0,1]$ корней уравнения $N(x,n) = c_{n-1}$, т. е. нужно, по сути дела, решить уравнение

$$\begin{aligned} 2 = \left(x + \frac{x^2}{2} + \frac{x^3}{3} + \ldots + \frac{x^{n-1}}{n-1}\right) + \\ + \left((1-x) + \frac{(1-x)^2}{2} + \frac{(1-x)^3}{3} + \ldots + \frac{(1-x)^{n-1}}{n-1}\right). \end{aligned} \quad (7)$$

Нетрудно убедиться, что при $n = 1, 2, 3, 4$ таких корней на интервале $[0,1]$ нет, а для каждого $n \geqslant 5$ существуют два таких корня $y(n)$ и $1 - y(n)$, где $0 < y(n) < \frac{1}{2} < (1 - y(n)) < 1$.

Приводим для некоторых $n \geqslant 5$ значения (с точностью до 4 знаков после запятой) этих корней (см. табл. 1).

Таблица 1

n	$y(n)$	$1-y(n)$
5	0,0287	0,9713
6	0,0791	0,9209
7	0,1068	0,8932
8	0,1236	0,8764
9	0,1344	0,8656
10	0,1416	0,8584

Таким образом, если имеем лишь один, два, три или четыре элемента в последовательности $U_1, U_2, \ldots, U_n$, то в любом случае надо отсчет рекордов начинать с первой случайной величины U_1, какое бы значение она ни принимала. Если через $e(n)$ обозначить математическое ожидание суммарного числа верхних и нижних рекордов среди величин $U_1, U_2, \ldots, U_n$, получаемое в результате описываемой оптимальной процедуры, то справедливы следующие равенства:

$$e(1) = c_1 = 2, \quad e(2) = c_2 = 3, \quad e(3) = c_3 = 3\frac{2}{4}, \quad e(4) = c_4 = 4\frac{1}{6}. \tag{8}$$

Если $n \geqslant 5$, предлагаемая процедура позволяет получить неравенства $\delta(n) > 0$, где

$$\delta(n) = e(n) - c(n). \tag{9}$$

При каждом фиксированном $n = 5, 6, \ldots$, имея набор с. в. $U_1, U_2, \ldots, U_n$, нужно получить наблюдаемое значение u_1 с. в. U_1. Если это значение лежит в интервале $[y(n), 1-y(n)]$, можно его принять за первую рекордную величину. Если $u_1 < y(n)$ или $u_1 > 1 - y(n)$, прощаемся с U_1 и переходим к случайной величине U_2. Это означает, что если $n-1 = 4$, уже просто рассматриваем рекорды в наборе $U_2, \ldots, U_5$. Если $n - 1 > 4$, наблюдаем значение u_2 с. в. U_2 и интересуемся, попадает ли оно в интервал $[y(n-1), 1-y(n-1)]$. Если такое событие происходит, начинаем отсчет рекордов с этого значения u_2. В противном случае продолжаем процедуру, переходя к с. в. U_3, и так далее. Несложно получить рекуррентное соотношение для величин $e(n), n = 5, 6, \ldots$ Действительно, имеют место равенства

$$\begin{aligned} 2y(n)e(n-1) + (1-2y(n))(1+c_{n-1}) - 2\int_{y(n)}^{1-y(n)} \left(\frac{u^2}{2} + \frac{u^3}{3} + \ldots + \frac{u^{n-1}}{n-1}\right) du = \\ = 2y(n)e(n-1) + 1 + c_{n-1} - 2y(n) - 2y(n)c_{n-1} + \\ + 2\left(\frac{y^3(n)}{6} + \frac{y^4(n)}{12} + \ldots + \frac{y^n(n)}{n(n-1)}\right) - \\ - 2\left(\frac{(1-y(n))^3}{6} + \frac{(1-y(n))^4}{12} + \ldots + \frac{(1-y(n))^n}{n(n-1)}\right). \end{aligned} \tag{10}$$

В частности, зная $e(4) = c_4 = 4\frac{1}{6}$ и $c_5 = 4\frac{17}{30}$, получаем, например, равенства

$$\begin{aligned} e(5) = 2y(5)e(4) - 2y(5) - 2c_4y(5) + 1 + c_4 + 2\left(\frac{y^3(5)}{6} + \frac{y^4(5)}{12} + \frac{y^5(5)}{20}\right) - \\ - 2\left(\frac{(1-y(5))^3}{6} + \frac{(1-y(5))^4}{12} + \frac{(1-y(5))^5}{20}\right) = 4{,}5690\ldots \end{aligned} \tag{11}$$

и

$$e(6) = 2y(6)e(5) - 2y(6) - 2c_5 y(5) + 1 + c_5 + 2\left(\frac{y^3(6)}{6} + \frac{y^4(6)}{12} + \frac{y^5(6)}{20} + \frac{y^6(6)}{30}\right) - \\ - 2\left(\frac{(1-y(6))^3}{6} + \frac{(1-y(6))^4}{12} + \frac{(1-y(6))^5}{20} + \frac{(1-y(6))^6}{30}\right) = 4{,}9219\ldots \quad (12)$$

Естественный интерес представляют величины «выигрышей» $\delta(n) = e(n) - c(n)$, $n = 5, 6, \ldots$, получаемых в результате применения рассматриваемой оптимизационной процедуры. Видим, что

$$\delta(n) = 2y(n)e(n-1) + 1 + c_{n-1} - 2y(n) - 2y(n)c_{n-1} + \\ + 2\left(\frac{y^3(n)}{6} + \frac{y^4(n)}{12} + \ldots + \frac{y^n(n)}{n(n-1)}\right) - \\ - 2\left(\frac{(1-y(n))^3}{6} + \frac{(1-y(n))^4}{12} + \ldots + \frac{(1-y(n))^n}{n(n-1)}\right) - c_n \quad (13)$$

удовлетворяет рекуррентному соотношению

$$\delta(n) = 2y(n)\delta(n-1) + 1 - \frac{2}{n} - 2y(n) + 2\left(\frac{y^3(n)}{6} + \frac{y^4(n)}{12} + \ldots + \frac{y^n(n)}{n(n-1)}\right) - \\ - 2\left(\frac{(1-y(n))^3}{6} + \frac{(1-y(n))^4}{12} + \ldots + \frac{(1-y(n))^n}{n(n-1)}\right), \quad n = 5, 6, \ldots \quad (14)$$

Для $n = 1, 2, 3$ и 4, как уже отмечалось, «выигрыша» нет, так как уравнение (7) не имеет корней на интервале $[0, 1]$, т. е. $\delta(1) = \delta(2) = \delta(3) = \delta(4) = 0$. Сравнительно просто находится значение

$$\delta(5) = 1 - \frac{2}{5} - 2y(5) + 2\left(\frac{y^3(5)}{6} + \frac{y^4(5)}{12} + \frac{y^5(5)}{20}\right) - \\ - 2\left(\frac{(1-y(5))^3}{6} + \frac{(1-y(5))^4}{12} + \frac{(1-y(5))^5}{20}\right) = 0{,}0024\ldots \quad (15)$$

Последующие вычисления величин $\delta(6), \delta(7), \ldots$ требуют более громоздких формул. Приведем некоторые из этих значений (см. табл. 2).

Таблица 2

n	c_n	$e(n)$	$\delta(n)$
6	4,9000	4,9219	0,0219
7	5,1857	5,2356	0,0499
8	5,4157	5,5156	0,0798
9	5,6579	5,7667	0,1087
10	5,8579	5,9932	0,1353

Величины $\delta(5), \delta(6), \ldots$ представляют собой элементы монотонно возрастающей последовательности. Интересно найти соответствующий предел

$$\delta^* = \lim \delta(n), \quad n \to \infty.$$

Для этого нужно найти вначале другой предел

$$y^* = \lim y(n), \quad n \to \infty,$$

где $y(n)$ — элементы монотонно возрастающей последовательности и $y(n)$ представляет собой меньший из двух корней уравнения (7), лежащих внутри интервала $[0,1]$. Поскольку

$$\sum_{k=1}^{\infty}\frac{x^k}{k} = -\ln(1-x), \tag{16}$$

получаем, что $y^* \to \infty$ является меньшим из двух корней уравнения

$$2 = -\ln x - \ln(1-x),$$

которое сводится к равенству

$$x(1-x) = \frac{1}{e^2}. \tag{17}$$

Находим, что нужный корень этого уравнения имеет вид

$$y = \frac{1}{2} - \frac{\sqrt{e^2-4}}{2e} = 0{,}161378209851\ldots \tag{18}$$

Предел при n выражения в правой части равенства (14) совпадает с величиной

$$2y^*\delta^* + 2y^* + 2(y^*-1)\left(\sum_{k=1}^{\infty}\frac{(y^*)^k}{k} - y^*\right),$$

т. е. этот предел равен $2y^*\delta^* + 2y^* + 2(y^*-1)(-\ln(1-y^*) - y^*)$. Переходя к пределу в левой и правой частях (14), получаем, что интересующее нас значение δ^* задается формулой

$$\delta^* = 2y^*\delta^* + 2y^* + 2(y^*-1)(-\ln(1-y^*) - y^*). \tag{19}$$

Из (19) следует равенство

$$\delta^* = \frac{2y^* + 2(y^*-1)(-\ln(1-y^*) - y^*)}{1-2y^*} = 0{,}440372915\ldots \tag{20}$$

Уже отмечалось, что соответствующие результаты, связанные с оптимальным выбором начала отсчета в наборах с. в. $X_1, X_2, \ldots, X_n$, обеспечивающим максимальное значение только математического числа верхних рекордов, были получены в работе [5]. Интересно сравнить «выигрыши» $\delta(n)$, приведенные выше, и соответствующие «выигрыши» $d(n)$ в процедурах, позволяющих максимизировать только лишь среднее число верхних рекордов. В табл. 3 приводятся некоторые значения величин $\delta(n)$ и $d(n)$ с точностью до 4 знаков после запятой.

Таблица 3

n	$\delta(n)$	$d(n)$
3	0	0,0654
4	0	0,1501
5	0,0024	0,2247
6	0,0219	0,2837
7	0,0499	0,3289
8	0,0798	0,3634
9	0,1087	0,3901
10	0,1353	0,4112

Отметим также, что предельные при $n \to \infty$ значения δ^* и d^* величин $\delta(n)$ и $d(n)$ даются, соответственно, равенствами $\delta^* = 0{,}44037\ldots$ и $d^* = 0{,}58197\ldots$

Литература

1. *Arnold B. C., Balakrishnan N., Nagaraja H. N.* Records. New York: John Wiley & Sons, 1998.
2. *Ahsanullah M., Nevzorov V. B.* Records via probability theory. Atlantis Press, 2015.
3. *Невзоров В. Б.* Рекорды. Математическая теория. М.: Фазис, 2000.
4. *Ahsanullah M., Yanev G. P.* Records and Branching Processes. New York: Nova Science Publishers, 2008.
5. *Nagaraja H. N.* On the expected values of record values // Australian Journal of Statistics. 1978. Vol. 20. P. 176–182.
6. *Rényi A.* Théorie des éléments saillants d'une suite d'observations // Colloquium on Combinatorial Methods in Probability Theory, Mathematisk Institut, Aarhus University, Aarhus, Denmark. 1962. P. 104–117.
7. *Shorrock R. W.* On record values and record times // Journal of Applied Probability. 1972. Vol. 9. P. 316–326.
8. *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.
9. *Дынкин Е. Б.* Оптимальный выбор момента остановки марковского процесса // ДАН СССР. 1963. Т. 150, № 2. С. 238–240.
10. *Невзоров В. Б., Товмасян С. А.* О максимальном значении среднего числа рекордов // Вестник С.-Петерб. ун-та. Сер. 1. 2014. Т. 1(59). Вып. 2. С. 196–200.

Статья поступила в редакцию 15 июля 2017 г.; рекомендована в печать 21 сентября 2017 г.

Контактная информация:

Бельков Игорь Владимирович — аспирант; igor.belkov@gmail.com
Невзоров Валерий Борисович — д-р физ.-мат. наук, проф.; vanev@mail.ru

On one problem of the optimal choice of record values

I. V. Belkov, V. B. Nevzorov

St. Petersburg State University, Universitetskaya nab., 7–9,
St. Petersburg, 199034, Russian Federation

For citation: Belkov I. V., Nevzorov V. B. On one problem of the optimal choice of record values. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy*, 2018, vol. 5 (63), issue 2, pp. 179–188. https://doi.org/10.21638/11701/spbu01.2018.201

Let independent random variables $X_1, X_2, \ldots$ have the same continuous distribution function. The upper record values $X(1) = X_1 < X(2) < \ldots$, generated by this sequence of variables, as well as the lower record values $x(1) = X_1 > x(2) > \ldots$ are considered. It is known that in this situation the mean value $c(n)$ of the sum of numbers of the both types of records among the first n of X's is given by equality $c(n) = 2(1 + \frac{1}{2} + \ldots + \frac{1}{n})$, $n = 1, 2, \ldots$. The problem considered here is how to obtain the maximal mean value $e(n)$ of the considered number of records among the rest of the random variables getting sequentially the observed values $x_1, x_2, \ldots$ of X's and selecting one of them as the initial point. It is not possible to come back to rejected elements of the sequence. Some procedures of the optimal choice of the initial element X_r are discussed. The corresponding tables for values $e(n)$ and differences $\delta(n) = e(n) - c(n)$ are presented for different values of n. The value of $\delta = \lim_{n\to\infty} \delta(n)$ is given also. In some sense the considered problem and presented in the paper optimizing procedures are very close to the classical "secretary problem" in which the probability to select the last record value in the set of independent identically distributed X's is maximized.

Keywords: record times, record values, expected number of records, uniform distribution, optimal choice problem.

References

1. Arnold B. C., Balakrishnan N., Nagaraja H. N., *Records* (John Wiley & Sons, New York, 1998).
2. Ahsanullah M., Nevzorov V. B., *Records via probability theory* (Atlantis Press, 2015, 255 p.)
3. Nevzorov V. B., *Records. Mathematical Theory.* In Ser. *Translations of Mathematical Monographs* (American Math. Society, **194**, 2001, 165 p.).
4. Ahsanullah M., Yanev G. P., *Records and Branching Processes* (Nova Science Publishers, New York, 2008).
5. Nagaraja H. N., "On the expected values of record values", *Australian Journal of Statistics* **20**, 176–182 (1978).
6. Rényi A., "Théorie des éléments saillants d'une suite d'observations", *Colloquium on Combinatorial Methods in Probability Theory, Mathematisk Institut, Aarhus University, Aarhus, Denmark*, 104–117 (1962).
7. Shorrock R. W., "On record values and record times", *Journal of Applied Probability* **9**, 316–326 (1972).
8. Yang M. C. K., "On the distribution of the inter-record times in an increasing population", *Journal of Applied Probability* **12**, 148–154 (1981).
9. Dynkin E. B., "Optimal choice of the stopping time of a Markov process", *Soviet Math.* **4**, 627–629 (1963).
10. Nevzorov V. B., Tovmasyan S. A. "On the maximal value of expectation of record numbers", *Vestnik St. Petersburg University: Mathematics* **47**, issue 2, 64–67 (2014).

Author's information:

Igor V. Belkov — igor.belkov@gmail.ru
Valery B. Nevzorov — vanev@mail.ru