УДК 519.6 Вестник СПбГУ. Математика. Механика. Астрономия. 2023. Т. 10 (68). Вып. 4
MSC 68U20, 65C60

Генерирование рекордов, получаемых из последовательностей независимых неодинаково распределенных случайных величин*

С. А. Петухов, А. В. Степанов

Балтийский федеральный университет имени И. Канта,
Российская Федерация, 236041, Калининград, ул. А. Невского, 14

Для цитирования: *Петухов С. А., Степанов А. В.* Генерирование рекордов, получаемых из последовательностей независимых неодинаково распределенных случайных величин // Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия. 2023. Т. 10 (68). Вып. 4. С. 736–748. https://doi.org/10.21638/spbu01.2023.410

В настоящей статье предлагаются алгоритмы генерирования рекордных моментов и величин, получаемых из последовательностей независимых неодинаково распределенных случайных величин, функции распределения которых заданы на одном носителе. Во введении статьи приводятся известные на данный момент методы и алгоритмы генерирования рекордных моментов и величин для случая, когда исходные случайные величины независимы и одинаково распределены; дается краткий обзор соответствующей литературы и указывается, что эффективные алгоритмы генерирования рекордов в случае независимых одинаково распределенных случайных величин основаны на свойстве марковости рекордов. Во втором разделе статьи выводятся распределения рекордных моментов и величин, получаемых из последовательностей независимых неодинаково распределенных случайных величин. Здесь также предлагаются алгоритмы генерирования рекордов для этого случая. Данные алгоритмы используют только что выведенные функции распределения рекордных моментов и величин и свойство марковости рекордов, которое справедливо и в случае, когда исходные случайные величины имеют разные функции распределения. В завершение работы (раздел 3), предложенные выше алгоритмы генерирования рекордных моментов и величин тестируются в экспериментах статистического моделирования: рекорды генерируются для последовательностей неодинаково распределенных случайных величин, имеющих функции распределения Гумбеля.

Ключевые слова: рекорды, функция распределения Гумбеля, метод обратных преобразований, метод выборки с отклонением, алгоритмы генерирования, время работы программы.

1. Введение. Развитие теории рекордов является актуальным и востребованным в связи с различными приложениями, возникающими в метеорологии, гидрологии, в страховом и финансовом бизнесе. Перепады температур и атмосферного давления, паводки рек, спортивные достижения, страховые и финансовые риски, различные модели, связанные со временем обслуживания, коррозией металлов, сопротивлением материалов, — все это и многое другое прекрасно описывается мате-

*Работа А. В. Степанова поддержана Министерством науки и высшего образования РФ (номер соглашения 075-02-2023-934).

матическим аппаратом этой теории. Более подробную информацию по этой тематике можно найти в книгах [1–3].

Интересным и современным направлением развития этой теории является разработка алгоритмов генерирования рекордных моментов и величин, которые были разработаны и реализованы в работах [4–14] для случая, когда исходные случайные величины независимы и имеют одинаковые функции распределения. В данной работе мы впервые разрабатываем алгоритмы генерирования рекордных моментов и величин для случая, когда исходные случайные величины независимы, но имеют разные функции распределения.

Пусть $X_1, X_2, \ldots$ — последовательность случайных величин. Определим последовательности рекордных моментов $L(n)$ и рекордных величин $X(n)$ следующим образом:

$$L(1) = 1,$$
$$L(n+1) = \min\left\{j : j > L(n),\, X_j > X_{L(n)}\right\}, \quad X(n) = X_{L(n)} \quad \text{для } n \geqslant 1.$$

Как уже отмечалось выше, существует несколько алгоритмов генерирования рекордов. Первый и наиболее простой алгоритм генерирования основан на прямом методе.

Прямой метод генерирования рекордных моментов и величин. *Генерируем случайную величину $X(1) = X_1$. Также положим $L(1) = 1$. Для дальнейшего генерирования последовательностей рекордных моментов и величин применим рекуррентный подход, который предполагает, что случайные величины $X(n)$ и $L(n)$ $(n \geqslant 1)$ уже сгенерированы. Продолжим генерировать величины X_i $(i > L(n))$ до тех пор, пока одна из них, скажем X_j, не станет больше, чем $X(n)$. Тогда положим $X(n+1) = X_j$ и $L(n+1) = j$.*

Прямой метод генерирования рекордов позволяет генерировать рекорды для любых распределений как в случае одинаково распределенных исходных случайных величин X_i $(i \geq 1)$, так и в случае, когда эти величины имеют разные функции распределения. Отметим серьезный недостаток этого метода. Если нам потребуется сгенерировать «большую» последовательность рекордов, то алгоритм, основанный на прямом методе, становится медленным и ресурсозатратным.

Ниже приведем известные эффективные алгоритмы генерирования рекордных моментов и величин для случая, когда исходные величины X_i $(i \geq 1)$ независимы и одинаково распределены. Отметим, что эти алгоритмы, предложенные в работах [4–8] и [10–14], используют свойство марковости рекордных моментов и величин. Перед этим кратко изложим основную информацию (см., например, лекции 15–17 книги [2]), касающуюся распределений рекордов в случае, когда исходные величины X_i $(i \geq 1)$ одинаково распределены.

Итак, пусть $X, X_1, X_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин с непрерывной функцией распределения $F(x) = P(X \leq x)$. Известно (см., например, [15]), что последовательность векторов $(L(n),\ X(n))$ $(n \geq 1)$ образует цепь Маркова, причем

$$P(L(n+1) = l_{n+1}, X(n+1) \leq x_{n+1} \mid L(n) = l_n, X(n) = x_n) = \\ = (F(x_{n+1}) - F(x_n))F^{l_{n+1}-l_n-1}(x_n), \quad (1)$$

где $l_{n+1} > l_n,\ x_{n+1} > x_n,\ n \geq 1$ и $l_1 = 1$. Из соотношения (1), в частности, следует, что

$$F_{X(n+1)|X(n)}(x_{n+1} \mid x_n) = P(X(n+1) \le x_{n+1} \mid X(n) = x_n) = \\ = \frac{F(x_{n+1}) - F(x_n)}{1 - F(x_n)} \quad (x_{n+1} > x_n). \quad (2)$$

Отметим, что последовательность рекордных величин $X(n)$ $(n \ge 1)$ также образует цепь Маркова. Покажем, что и последовательность рекордных моментов $L(n)$ $(n \ge 1)$ образует цепь Маркова с переходными вероятностями

$$P(L(n+1) = l_{n+1} \mid L(n) = l_n) = \frac{l_n}{(l_{n+1} - 1)l_{n+1}} \quad (l_{n+1} > l_n,\ n \ge 1). \quad (3)$$

Откуда следует, что

$$P(L(n+1) \le l_{n+1} \mid L(n) = l_n) = 1 - \frac{l_n}{l_{n+1}} \quad (l_{n+1} > l_n,\ n \ge 1). \quad (4)$$

Для эффективного генерирования последовательностей рекордных моментов и величин воспользуемся свойством марковости рекордов. Если обратная функция F^{-1} может быть найдена явно, то для генерирования рекордных величин $X(n)$ $(n \ge 1)$ можно применять метод обратных преобразований и рекуррентный подход, основанный на равенстве (2). Здесь значение $X(n+1)$ $(n \ge 1)$ полагается равным $F^{-1}_{X(n+1)|X(n)}(u \mid x_n)$, где $U = u$ — генерация случайного числа, т. е. U — равномерно распределенная на отрезке $[0, 1]$ случайная величина. Приведем соответствующий алгоритм генерирования (см., например, [4]).

Алгоритм 1.1. *Последовательность $X(n)$ $(n \geqslant 1)$ может быть сгенерирована следующим образом. Сгенерируем величину $X(1) = X_1 = x_1$. Для генерирования $X(n)$ $(n \ge 2)$ воспользуемся рекуррентным методом. Пусть величина $X(n) = x_n$ $(n \ge 1)$ уже сгенерирована. Генерируем случайное число $U = u$. Положим $X(n+1)$ равным*

$$F^{-1}(u(1 - F(x_n)) + F(x_n)).$$

Если же обратная функция F^{-1} не может быть найдена аналитически, то для генерирования рекордных величин можно применять либо таблицы обратной функции F^{-1}, либо метод выборки с отклонением. Приведем также алгоритм генерирования рекордных моментов, основанный на равенстве (4). Данный алгоритм был предложен в работе [10].

Алгоритм 1.2. *Последовательность $L(n)$ $(n \geqslant 1)$ может быть сгенерирована следующим образом. Положим $L(1) = 1$. Для генерирования $L(n)$ $(n \ge 2)$ воспользуемся рекуррентным методом. Пусть величина $L(n) = l_n$ $(l_n \ge n \ge 1)$ уже сгенерирована. Генерируем случайное число $U = u$. Найдем то единственное $l_{n+1} > l_n$ такое, что $u \in [l_n/l_{n+1}, l_n/(l_{n+1} - 1))$. Положим $L(n+1) = l_{n+1}$.*

Дальнейшее содержание нашей работы следующее. В разделе 2 работы приведем функции распределения рекордных моментов и величин для случая, когда исходные величины X_i $(i \ge 1)$ независимы и неодинаково распределены. Здесь же предложим алгоритм генерирования рекордных моментов и величин, использующий эти функции распределения. В разделе 3 статьи протестируем предложенные в разделе 2 алгоритмы генерирования с помощью экспериментов статистического моделирования. В данных экспериментах рекорды будут генерироваться для выборок, имеющих функции распределения Гумбеля.

2. Распределения рекордов и алгоритмы их генерирования для последовательностей независимых неодинаково распределенных случайных величин. Пусть $X_1, X_2, \ldots$ — последовательность независимых случайных величин, имеющих абсолютно непрерывные функции распределения $F_1, F_2, \ldots$ и плотности $f_1, f_2, \ldots$ Пусть данные функции распределения заданы на одном носителе. Пусть $f(x_1, l_1, x_2, l_2, \ldots, x_n, l_n)$, где $l_1 = 1$ — плотность-функция вероятности вектора $\bar{a} = (X(1), L(1), X(2), L(2), \ldots, X(n), L(n))$, где плотность описывает поведение непрерывных составляющих $X(1), X(2), \ldots, X(n)$ вектора $\bar{a}$, а функция вероятности задает поведение дискретных координат $L(1), L(2), \ldots, L(n)$ вектора $\bar{a}$. Несложно показать, что для $x_1 < x_2 < \ldots < x_n$ и $1 = l_1 < l_2 < \ldots < l_n$ данная плотность-функция вероятности имеет вид

$$f(x_1, l_1, x_2, l_2, \ldots, x_n, l_n) = \\ = f_{l_n}(x_n)\Pi_{j=1}^{n-1} f_{l_j}(x_j) F_{l_j+1}(x_j) F_{l_j+2}(x_j) \ldots F_{l_{j+1}-1}(x_j) \quad (n \geq 2). \quad (5)$$

Из равенства (5) следует, что последовательность векторов $(L(n),\ X(n))$ $(n \geq 1)$ образует цепь Маркова, причем

$$P(L(n+1) = l_{n+1}, X(n+1) \leq x_{n+1} \mid L(n) = l_n, X(n) = x_n) = \\ = F_{l_n+1}(x_n) \ldots F_{l_{n+1}-1}(x_n)(F_{l_{n+1}}(x_{n+1}) - F_{l_{n+1}}(x_n)), \quad (6)$$

где $l_{n+1} > l_n,\ x_{n+1} > x_n$ $(n \geq 1)$ и $F_{l_n+1}(x_n) \ldots F_{l_{n+1}-1}(x_n) = 1$ при $l_{n+1} = l_n + 1$. Отметим, что соотношение (6) обобщает известное соотношение (1). Из соотношения (6), в частности, следует, что

$$P(L(n+1) = l_{n+1} \mid L(n) = l_n, X(n) = x_n) = \\ = F_{l_n+1}(x_n) \ldots F_{l_{n+1}-1}(x_n)(1 - F_{l_{n+1}}(x_n)). \quad (7)$$

Откуда

$$P(L(n+1) \leq l_{n+1} \mid L(n) = l_n, X(n) = x_n) = \\ = \sum_{i=l_n+1}^{l_{n+1}} F_{l_n+1}(x_n) \ldots F_{i-1}(x_n)(1 - F_i(x_n)). \quad (8)$$

Из соотношений (6) и (7) в свою очередь вытекает, что

$$P(X(n+1) \leq x_{n+1} \mid L(n+1) = l_{n+1}, X(n) = x_n) = \\ = \frac{F_{l_{n+1}}(x_{n+1}) - F_{l_{n+1}}(x_n)}{1 - F_{l_{n+1}}(x_n)} \quad (l_{n+1} > l_n,\ x_{n+1} > x_n). \quad (9)$$

Пользуясь найденными выше условными функциями распределения, предложим алгоритм генерирования рекордных моментов и величин для случая, когда исходные величины X_i $(i \geq 1)$ независимы и неодинаково распределены.

Алгоритм 2.1. *Последовательность* $(L(n), X(n))$ $(n \geqslant 1)$ *может быть сгенерирована следующим образом. Положим* $L(1) = 1$ *и сгенерируем* $X(1) = X_1 = x_1$. *Для генерирования* $(L(n), X(n))$ $(n \geq 2)$ *воспользуемся рекуррентным методом. Пусть величины* $L(n) = l_n$ *и* $X(n) = x_n$ $(l_n \geq n \geq 1)$ *уже сгенерированы. Генерируем случайное число* $U = u$. *Для генерирования рекордного момента* $L(n+1)$ *воспользуемся соотношением (8). Найдем то единственное* $l_{n+1} > l_n$ *такое, что*

$$\sum_{i=l_n+1}^{l_{n+1}-1} F_{l_n+1}(x_n)\ldots F_{i-1}(x_n)(1-F_i(x_n)) \leq u < \sum_{i=l_n+1}^{l_{n+1}} F_{l_n+1}(x_n)\ldots F_{i-1}(x_n)(1-F_i(x_n)). \tag{10}$$

Положим $L(n+1) = l_{n+1}$. *Для генерирования* $X(n+1)$ *воспользуемся (9). Положим* $X(n+1)$ *равным*

$$F^{-1}_{l_{n+1}}(v(1 - F_{l_{n+1}}(x_n)) + F_{l_{n+1}}(x_n)),$$

где $V = v$ *— генерация случайного числа.*

Отметим, что алгоритм 2.1 позволяет генерировать последовательности рекордных величин $X(n)$ $(n \geq 1)$ только в случае, когда все функции распределения F_i $(i \geq 1)$ исходных случайных величин X_i $(i \geq 1)$ имеют обратные функции. Для генерирования рекордных величин в случае, когда функции распределения F_i $(i \geq 1)$ не имеют обратных функций, нужно разрабатывать другой алгоритм генерирования, основанный на методе выборки с отклонением.

3. Генерирование рекордов для последовательностей независимых случайных величин, имеющих функции распределения Гумбеля. Пусть X_i $(i \geq 1)$ — последовательность независимых случайных величин с функциями распределения

$$F_i(x) = e^{-ie^{-x}} \quad (x \in \mathbb{R})$$

и соответствующими плотностями

$$f_i(x) = ie^{-x}e^{-ie^{-x}} \ (x \in \mathbb{R}).$$

В этом случае равенство (5) сводится к равенству

$$f(x_1, l_1, x_2, l_2, \ldots, x_n, l_n) = l_n e^{-x_n} e^{-l_n e^{-x_n}} \Pi_{j=1}^{n-1} l_j e^{-x_j} e^{-(l_j+\ldots+l_{j+1}-1)e^{-x_j}}. \tag{11}$$

Как уже отмечалось в алгоритме 2.1 (см. (10)), генерирование величины $L(n+1)$ происходит при помощи известных значений $L(n) = l_n$ и $X(n) = x_n$. Отметим, что, если исходные функции распределения F_i $(i \geq 1)$ являются функциями распределений Гумбеля, генерировать рекордные моменты можно проще, используя в соответствующих алгоритмах генерирования безусловные функции распределения этих рекордных моментов. Интегрируя равенство (11) по переменным $x_1 < x_2 < \ldots < x_n$, находим функцию вероятности вектора $(L(1), L(2), \ldots, L(n))$. Она будет иметь вид

$$P(L(1) = 1,\ L(2) = l_2) = \frac{4}{(l_2-1)l_2(l_2+1)} \quad (l_2 \geq 2)$$

и

$$P(L(1) = 1, L(2) = l_2, \ldots, L(n) = l_n) = \\ = \frac{2^n}{(l_2-1)\ldots(l_{n-1}-1)(l_n-1)l_n(l_n+1)} \quad (n \geq 3).$$

Откуда заключаем, что последовательность $L(n)$ $(n \geq 1)$ образует цепь Маркова с переходными вероятностями

$$P(L(n) = l_n \mid L(n-1) = l_{n-1}) = \frac{2l_{n-1}(l_{n-1}+1)}{(l_n-1)l_n(l_n+1)} \quad (l_n > l_{n-1},\ n \geq 2).$$

Находим условную функцию распределения $L(n)$:

$$P(L(n) \leq l_n \mid L(n-1) = l_{n-1}) = 1 - \frac{l_{n-1}(l_{n-1}+1)}{l_n(l_n+1)} \quad (l_n > l_{n-1},\ n \geq 2) \tag{12}$$

и условное математическое ожидание $L(n)+1$:

$$E(L(n)+1 \mid L(n-1) = l_{n-1}) = 2(l_{n-1}+1) \quad (n \geq 2).$$

Пользуясь свойствами условного математического ожидания и последним равенством, получаем рекуррентное соотношение

$$E(L(n)+1) = 2E(L(n-1)+1) \quad (n \geq 2).$$

Откуда следует, что

$$EL(n) = 2^n - 1 \quad (n \geq 1).$$

Здесь также можно показать, что $DL(n) = \infty$ $(n \geq 2)$. Определим функцию $g_m(n)$ равенствами

$$g_0(n) = 1 (n \geq 1), \quad g_1(n) = \sum_{i=1}^{n} \frac{1}{i}\ (n \geq 1), \quad g_m(n) = \sum_{i=m}^{n} \frac{g_{m-1}(i-1)}{i} \quad (n \geq m \geq 1).$$

Пользуясь соотношением (11), находим, что распределение $X(n)$ имеет вид

$$P(X(n) \leq x_n) = 2^n \sum_{k=n-1}^{\infty} \frac{g_{n-2}(k-1)}{k(k+1)(k+2)} \cdot e^{-\frac{(k+1)(k+2)}{2}e^{-x_n}}. \tag{13}$$

Известно, что $EX_1 = \gamma$, где $\gamma = 0.577216\ldots$ — константа Эйлера. Пользуясь этим равенством, несложно показать, что

$$\int_{\mathbb{R}} x\alpha e^{-x} e^{-\alpha e^{-x}} dx = \gamma + \ln(\alpha) \quad (\alpha > 0). \tag{14}$$

Из равенств (13) и (14), в частности, следует, что

$$EX(n) = \gamma + 2^n \sum_{k=n-1}^{\infty} \frac{g_{n-2}(k-1)}{k(k+1)(k+2)} \cdot \ln\left(\frac{(k+1)(k+2)}{2}\right) \quad (n \geq 2).$$

Результаты численного подсчета $EX(n)$ для $n = 2, 4, 6, 8, 10$ представлены в табл. 1.

Таблица 1. **Численные значения математических ожиданий** $EX(n)$

$EX(2)$	$EX(4)$	$EX(6)$	$EX(8)$	$EX(10)$
2.107690	4.498777	6.634558	8.688925	10.711888

Пусть $U = u$ — генерация случайного числа такая, что

$$P(L(n+1) \leq l_{n+1} - 1 \mid L(n) = l_n) \leq u < P(L(n+1) \leq l_{n+1} \mid L(n) = l_n).$$

Пользуясь равенством (12), запишем последние два неравенства в виде

$$1 - \frac{l_n(l_n+1)}{(l_{n+1}-1)l_{n+1}} \leq u < 1 - \frac{l_n(l_n+1)}{l_{n+1}(l_{n+1}+1)}.$$

Отметим, что случайная величина $1 - U$, так же как и величина U, имеет стандартное равномерное распределение, и, следовательно, в алгоритмах генерирования величина $1 - u$ может быть заменена на u. Из последних двух неравенств получаем неравенства

$$\frac{-1 + \sqrt{1 + 4l_n(l_n+1)/u}}{2} < l_{n+1} \leq \frac{1 + \sqrt{1 + 4l_n(l_n+1)/u}}{2}. \tag{15}$$

Итак, теперь последовательность рекордных моментов может быть сгенерирована при помощи отдельного простого и эффективного алгоритма.

Алгоритм 3.1. *Последовательность $L(n)$ $(n \geqslant 1)$ может быть сгенерирована следующим образом. Положим $L(1) = 1$. Для генерирования $L(n)$ $(n \geq 2)$ воспользуемся рекуррентным методом. Пусть величина $L(n) = l_n$ $(l_n \geq n \geq 1)$ уже сгенерирована. Генерируем случайное число $U = u$. Для генерирования $L(n+1)$ воспользуемся соотношением (15). Положим $L(n+1)$ равным*

$$\left[\frac{1 + \sqrt{1 + 4l_n(l_n+1)/u}}{2}\right],$$

где $[x]$ — целая часть числа x.

К сожалению, последовательность рекордных величин генерируется чуть сложнее. В алгоритм их генерирования приходится также включать и генерирование самих рекордных моментов.

Алгоритм 3.2. *Последовательность $(L(n), X(n))$ $(n \geqslant 1)$ может быть сгенерирована следующим образом. Положим $L(1) = 1$ и сгенерируем $X(1) = X_1 = x_1$. Для генерирования $(L(n), X(n))$ $(n \geq 2)$ воспользуемся рекуррентным методом. Пусть величины $L(n) = l_n$ $(l_n \geq n \geq 1)$ и $X(n) = x_n$ $(n \geq 1)$ уже сгенерированы. Генерируем случайное число $U = u$. Для генерирования $L(n+1)$ воспользуемся соотношением (15). Положим $L(n+1)$ равным*

$$\left[\frac{1 + \sqrt{1 + 4l_n(l_n+1)/u}}{2}\right].$$

Для генерирования $X(n+1)$ воспользуемся (9). Положим $X(n+1)$ равным

$$-\ln\left(-\ln\left(e^{-l_{n+1}e^{-x_n}} + v\left(1 - e^{-l_{n+1}e^{-x_n}}\right)\right)\right) + \ln(l_{n+1}), \tag{16}$$

где $V = v$ — генерация случайного числа.

Для сравнения работы алгоритма генерирования рекордов, основанного на прямом методе, и алгоритма 3.2 мы провели следующие эксперименты статистического моделирования для «малых» значений n (≤ 10). Пользуясь прямым методом генерирования, мы сгенерировали сто тысяч раз рекордные моменты и величины $L(n)$, $X(n)$ ($n = 1, \ldots, 10$) и нашли средние значения $\bar{L}(n)$, $\bar{X}(n)$ для каждого $n = 1, \ldots, 10$. Найденные выборочные средние $\bar{X}(n), \bar{L}(n)$ ($n = 2,\ 4,\ 6,\ 8,\ 10$) и их отклонения от математических ожиданий $EX(n) - \bar{X}(n)$, $EL(n) - \bar{L}(n)$ ($n = 2, 4, 6, 8, 10$) представлены соответственно в табл. 2 и 3.

Таблица 2. **Выборочные средние $\bar{X}(n)$ и их отклонения от математических ожиданий $EX(n) - \bar{X}(n)$, найденные прямым методом**

$X(2)$	$X(4)$	$X(6)$	$X(8)$	$X(10)$
2.109112	4.471538	6.584491	8.630772	10.642049
$EX(2) - X(2)$	$EX(4) - X(4)$	$EX(6) - X(6)$	$EX(8) - X(8)$	$EX(10) - X(10)$
−0.001423	0.027239	0.050066	0.0581528	0.069839

Таблица 3. **Выборочные средние $\bar{L}(n)$ и их отклонения от математических ожиданий $EL(n) - \bar{L}(n)$, найденные прямым методом**

$\bar{L}(2)$	$\bar{L}(4)$	$\bar{L}(6)$	$\bar{L}(8)$	$\bar{L}(10)$
3.0179	13.1141	54.0447	200.1451	761.4643
$EL(2) - L(2)$	$EL(4) - L(4)$	$EL(6) - L(6)$	$EL(8) - L(8)$	$EL(10) - L(10)$
−0.0179	1.8859	8.9553	54.8549	261.5357

Из табл. 2 следует, что оценки $\bar{X}(n)$ математических ожиданий $EX(n)$ достаточно точны. В то же время табл. 3 показывает, что оценки $\bar{L}(n)$ сильно отличаются от $EL(n)$. Это происходит, по видимому, потому, что хотя $EL(n)$ ($n \geq 1$) существуют, но уже $DL(n) = \infty$ ($n \geq 2$) и, следовательно, значения рекордных моментов сильно разбросаны относительно своих математических ожиданий. Время работы программы составило 97.49 с. Генерирование проходило в среде MatLab. Использовался компьютер Intel(R) Core(TM) i3-9100 CPU 3.60GHz 8 Gb.

Далее, пользуясь алгоритмом 3.2, мы сгенерировали сто тысяч раз десять первых рекордных моментов и величин. Полученные средние величины $\bar{X}(n)$, $\bar{L}(n)$ ($n = 2, 4, 6, 8, 10$) и отклонения средних от математических ожиданий $EX(n) - \bar{X}(n)$, $EL(n) - \bar{X}(n)$ ($n = 2, 4, 6, 8, 10$) представлены соответственно в табл. 4 и 5.

Таблица 4. **Выборочные средние $\bar{X}(n)$ и их отклонения от математических ожиданий $EX(n) - \bar{X}(n)$, найденные при помощи алгоритма 3.2**

$X(2)$	$X(4)$	$X(6)$	$X(8)$	$X(10)$
2.197364	4.729198	6.951961	9.065495	11.132919
$EX(2) - X(2)$	$EX(4) - X(4)$	$EX(6) - X(6)$	$EX(8) - X(8)$	$EX(10) - X(10)$
−0.089674	−0.230421	−0.317403	−0.376570	−0.421031

Таблица 5. **Выборочные средние $\bar{L}(n)$ и их отклонения от математических ожиданий $EL(n) - \bar{L}(n)$, найденные при помощи алгоритма 3.2**

$L(2)$	$L(4)$	$L(6)$	$L(8)$	$L(10)$
3.0068	15.5333	66.0106	252.9800	1065.3576
$EL(2) - L(2)$	$EL(4) - L(4)$	$EL(6) - L(6)$	$EL(8) - L(8)$	$EL(10) - L(10)$
−0.0068	−0.5333	−3.0106	2.0200	−42.3576

Время работы программы составило 69.95 с. Сравнивая табл. 2 и 3 и табл. 4 и 5, делаем вывод, что при генерировании рекордных величин прямым методом для «малых» n (≤ 10) получаются более точные результаты по сравнению с генерированием, использующим свойство марковости рекордов. В то же время ситуация с генерированием рекордных моментов прямо противоположная — алгоритм генерирования, использующий свойство марковости рекордов, дает гораздо более точные результаты.

Как уже отмечалось выше, алгоритм, основанный на прямом методе генерирования рекордов, работает медленно при больших значениях n. Так, для генерирования прямым методом 20 рекордов десять тысяч раз потребовалось время 1471.42 с. Пользуясь же алгоритмом 3.2, мы сгенерировали десять тысяч раз рекордные величины $X(1), \ldots, X(20)$ за 0.51 с. Среднее значение по десяти тысячам получилось $\bar{X}(20) = 21.545271$.

В свою очередь, работу алгоритма 3.2 при $n > 20$ осложняет то обстоятельство, что выражение $l_{n+1}e^{-x_n}$ принимает очень близкие к нулю значения. В результате работы программы значения $X(n)$ ($n \geq 40$) выдаются как бесконечные. Для устранения этого технического недостатка мы доработаем алгоритм 3.2. Воспользуемся методом выборки с отклонением.

Метод выборки с отклонением. Предположим, что мы можем генерировать случайную величину Y с плотностью распределения $g(y) = G'(y)$ (назовем ее доминирующей плотностью) методом обратного преобразования. Предположим также, что генерирование случайной величины X с плотностью распределения $f(x) = F'(x)$ методом обратного преобразования невозможно или крайне сложно (как в случае генерирования рекордов, при $n > 20$). Пусть распределения F и G имеют один носитель. Определим константу c равенством:

$$c = \sup_x \frac{f(x)}{g(x)}.$$

Генерирование случайной величины X осуществляется следующим образом.

Шаг 1. Генерируем значение $Y = y$ с плотностью g.

Шаг 2. Генерируем случайное число $U = u$.

Шаг 3. Если $u < \frac{f(y)}{cg(y)}$, то полагаем $X = y$. В противном случае, возвращаемся к шагу 1.

Подбор подходящей случайной величины Y (и связанной с ней константы $c > 1$) происходит таким образом, чтобы отношение $\frac{f(y)}{cg(y)}$ было близко к единице при всех значениях y. Тогда условие $u < \frac{f(y)}{cg(y)}$ на шаге 3 будет выполняться часто и число итераций для успешного генерирования одного значения случайной величины X становится минимальным, а алгоритм, соответственно, эффективным.

Из (9) следует, что условная плотность величины $X(n+1)$ в случае выборок с функциями распределений экстремальных значений имеет вид

$$f_{X(n+1)|X(n),L(n+1)}(x_{n+1} \mid x_n, l_{n+1}) = \frac{l_{n+1}e^{-x_{n+1}}e^{-l_{n+1}e^{-x_{n+1}}}}{1 - e^{-l_{n+1}e^{-x_n}}} \quad (x_{n+1} > x_n).$$

В качестве доминирующей плотности выберем плотность

$$g(x_{n+1} \mid x_n) = e^{-x_{n+1}+x_n} \quad (x_{n+1} > x_n).$$

Найдем константу c:

$$c = \sup_{x_{n+1} > x_n} \frac{f_{X(n+1)|X(n),L(n+1)}(x_{n+1} \mid x_n, l_{n+1})}{g(x_{n+1} \mid x_n)} = \frac{l_{n+1} e^{-x_n}}{1 - e^{-l_{n+1} e^{-x_n}}}.$$

Таким образом, имеем

$$\frac{f_{X(n+1)|X(n),L(n+1)}(x_{n+1} \mid x_n, l_{n+1})}{cg(x_{n+1} \mid x_n)} = e^{-l_{n+1} e^{-x_{n+1}}}.$$

Как мы уже отмечали выше, выражение $e^{-l_{n+1}e^{-x_{n+1}}}$ при $n > 20$ принимает значения близкие к единице, и, следовательно, неравенство $u < e^{-l_{n+1}e^{-x_{n+1}}}$ будет выполнятся почти всегда. Таким образом, алгоритм генерирования рекордных величин, основанный при $n > 20$ на методе выборки с отклонением, будет эффективным. Приведем окончательный алгоритм генерирования рекордов.

Алгоритм 3.3. *Последовательность $(L(n), X(n))$ $(n \geqslant 1)$ может быть сгенерирована следующим образом. Положим $L(1) = 1$ и сгенерируем $X(1) = X_1 = x_1$. Для генерирования $(L(n), X(n))$ при $2 \leq n \leq 20$ воспользуемся рекуррентным методом. Пусть величины $L(n) = l_n$ $(l_n \geq n \geq 1)$ и $X(n) = x_n$ $(n \geq 1)$ уже сгенерированы. Генерируем случайное число $U = u$. Для генерирования $L(n+1)$ воспользуемся соотношением (15). Положим $L(n+1)$ равным*

$$\left[\frac{1 + \sqrt{1 + 4l_n(l_n+1)/u}}{2}\right].$$

Для генерирования $X(n+1)$ воспользуемся (9). Положим $X(n+1)$ равным

$$-\ln\left(-\ln\left(e^{-l_{n+1}e^{-x_n}} + v\left(1 - e^{-l_{n+1}e^{-x_n}}\right)\right)\right) + \ln(l_{n+1}),$$

где $V = v$ — генерация случайного числа.

Для генерирования $(L(n), X(n))$ при $n \geq 21$ также воспользуемся рекуррентным методом. Пусть величины $L(n) = l_n$ $(l_n \geq n \geq 1)$ и $X(n) = x_n$ $(n \geq 20)$ уже сгенерированы. Генерируем случайное число $U = u$. Для генерирования $L(n+1)$ воспользуемся соотношением (15). Положим $L(n+1)$ равным

$$\left[\frac{1 + \sqrt{1 + 4l_n(l_n+1)/u}}{2}\right].$$

Шаг 1. Сгенерируем случайное число $V = v$. Положим $Y = y$ равным $x_n - \ln(v)$.

Шаг 2. Генерируем случайное число $R = r$.

Шаг 3. Если $r < e^{-l_{n+1}e^{-x_n}}$, то положим $X(n+1) = y$. В противном случае, вернемся к шагу 1.

Пользуясь алгоритмом 3.3, мы сгенерировали сто тысяч раз рекордные величины $X(1), \ldots, X(100)$ и нашли соответствующие средние значения, которые представлены в табл. 6 для $n = 20, 40, 60, 80, 100$.

Время работы программы составило 1150.88 с.

В завершение работы отметим, что мы отдельно протестировали работу алгоритма 3.1. С помощью этого алгоритма мы сгенерировали тысячу раз рекордные

Таблица 6. **Выборочные средние $\bar{X}(n)$, найденные при помощи алгоритма 3.3**

$X(20)$	$X(40)$	$X(60)$	$X(80)$	$X(100)$
21.251638	40.581788	60.144273	79.717342	99.318321

моменты $L(2), \ldots, L(500)$ и нашли соответствующие средние величины для каждого момента. Приведем только последнее среднее значение $\bar{L}(500) = 1.205885447804962 * 10^{145}$. Время работы программы составило всего 0.61 с.

Литература

1. Arnold B. C., Balakrishnan N., Nagaraja H. N. *Records.* John Wiley & Sons (1998).
2. Невзоров В. Б. *Рекорды. Математическая теория.* Москва, Фазис (2000).
3. Ahsanullah M., Nevzorov V. B. *Records via Probability Theory.* Atlantis Press (2015).
4. Bairamov I., Stepanov A. Numbers of near bivariate record-concomitant observations. *Journal of Multivariate Analysis* **102**, 908–917 (2011).
5. Luckett D. J. *Statistical Inference Based on Upper Record Values.* PhD thesis. The College of William and Mary (2013).
6. Nevzorov V. B., Stepanov A. Records with confirmation. *Statist. Probab. Lett.* **95**, 39–47 (2014).
7. Pakhteev A., Stepanov A. Simulation of Gamma Records. *Statist. Probab. Lett.* **119**, 204–212 (2016).
8. Stepanov A., Berred A., Nevzorov V. B. Concomitants of records: Limit results, generation techniques, correlation. *Statistics & Probability Letters* **109**, 184–188 (2016).
9. Balakrishnan N., So H. Y., Zhu X. J. On Box-Muller Transformation and Simulation of Normal Record Data. *Communication in Statistics — Simulation and Computations* **45**, 3670–3682 (2016).
10. Пахтеев А. И., Степанов А. В. Генерирование больших последовательностей нормальных рекордных величин и максимумов. *Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия* **5** (63), вып. 3, 431–440 (2018).
11. Pakhteev A., Stepanov A. On Simulation of Normal Records. *Communication in Statistics — Simulation and Computation* **48** (8), 2413–2424 (2019).
12. Pakhteev A., Stepanov A. Discrete records: Limit theorems for their spacings and generation methods. *Statistics and Probability Letters* **148**, 134–142 (2019).
13. Stepanov A. On simulation of weak records. *Communication in Statistics — Simulation and Computation* **48** (3), 797–806 (2019).
14. Stepanov A. Simulation of concomitants of records. Communication in Statistics — Simulation and Computation (2022). https://doi.org/10.1080/03610918.2022.213716
15. Stepanov A. Conditional moments of record times. *Statist. Pap.* **44** (1), 131–140 (2003).

Статья поступила в редакцию 23 января 2023 г.;
доработана 5 февраля 2023 г.;
рекомендована к печати 18 мая 2023 г.

Контактная информация:

Петухов Сергей Алексеевич — студент; lexuslfa477@gmail.com
Степанов Алексей Васильевич — д-р физ.-мат. наук, проф.; alexeistep45@mail.ru

Generation of records obtained from sequences of independent and non-identically distributed random variables*

S. A. Petukhov, A. V. Stepanov

Immanuel Kant Baltic Federal University, 14, ul. A. Nevskogo, Kaliningrad, 236041, Russian Federation

For citation: Petukhov S. A., Stepanov A. V. Generation of records obtained from sequences of independent and non-identically distributed random variables. *Vestnik of Saint Petersburg*

*Stepanov's A. V. work was supported by the Ministry of Science and Higher Education of the Russian Federation (agreement number 075-02-2023-934).

University. Mathematics. Mechanics. Astronomy, 2023, vol. 10 (68), issue 4, pp. 736–748.
https://doi.org/10.21638/spbu01.2023.410 (In Russian)

Generation algorithms of record times and values obtained from sequences of independent and non-identically distributed random variables which distribution functions are defined on a common support are proposed in the present paper. Known algorithms of generation of record times and values are given in introduction for the case when the initial random variables are independent and identically distributed. The brief review of scientific literature associated with this topic is also given in Introduction. It is also pointed out there that all efficient algorithms of record generation are based on the Markov property of records. In Section 2, the distribution functions of record times and values are derived for the case when the initial random variables are independent and non-identically distributed. The corresponding record generation algorithms are for the first time proposed. These algorithms are based on the derived distributions and the Markov property of records that also holds in the case when the initial observations are independent but non-identically distributed. In the end of this work, in Section 3, the proposed algorithms are tested by simulation experiments. In these experiments the records are generated for the case when the initial random variables have the Gumbel distribution functions.

Keywords: records, Gumbel distribution function, inverse-transform method, rejection method, generation algorithms, elapsed time.

References

1. Arnold B. C., Balakrishnan N., Nagaraja H. N. *Records.* John Wiley & Sons NY (1998).

2. Nevzorov V. *Records: Mathematical Theory.* American Mathematical Society, Providence, Rhode Island. (2001). [Rus. ed.: Nevzorov V. B. *Rekordy. Matematicheskaia teoriia.* Moscow, Fazis Publ. (2000)].

3. Ahsanullah M., Nevzorov V. B. *Records via Probability Theory.* Atlantis Press (2015).

4. Bairamov I., Stepanov A. Numbers of near bivariate record-concomitant observations. *Journal of Multivariate Analysis* **102**, 908–917 (2011).

5. Luckett D. J. *Statistical Inference Based on Upper Record Values.* PhD thesis. The College of William and Mary (2013).

6. Nevzorov V. B., Stepanov A. Records with confirmation. *Statist. Probab. Lett.* **95**, 39–47 (2014).

7. Pakhteev A., Stepanov A. Simulation of Gamma Records. *Statist. Probab. Lett.* **119**, 204–212 (2016).

8. Stepanov A., Berred A., Nevzorov V. B. Concomitants of records: Limit results, generation techniques, correlation. *Statistics & Probability Letters* **109**, 184–188 (2016).

9. Balakrishnan N., So H. Y., Zhu X. J. On Box-Muller Transformation and Simulation of Normal Record Data. *Communication in Statistics — Simulation and Computations* **45**, 3670–3682 (2016).

10. Pakhteev A., Stepanov A., 2018. Generating large sequences of normal maxima via record values. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy* **5** (63), iss. 3, 431–440 (2018). (In Russian) [Engl. transl.: *Vestnik St. Petersburg University, Mathematics* **51**, iss. 3, 260–266 (2018)].

11. Pakhteev A., Stepanov A. On Simulation of Normal Records. *Communication in Statistics — Simulation and Computation* **48** (8), 2413–2424 (2019).

12. Pakhteev A., Stepanov A. Discrete records: Limit theorems for their spacings and generation methods. *Statistics and Probability Letters* **148**, 134–142 (2019).

13. Stepanov A. On simulation of weak records. *Communication in Statistics — Simulation and Computation* **48** (3), 797–806 (2019).

14. Stepanov A. Simulation of concomitants of records. Communication in Statistics — Simulation and Computation (2022). https://doi.org/10.1080/03610918.2022.213716

15. Stepanov A. Conditional moments of record times. *Statist. Pap.* **44** (1), 131–140 (2003).

Received: January 23, 2023
Revised: February 5, 2023
Accepted: May 18, 2023

A u t h o r s’ i n f o r m a t i o n:

Sergei A. Petukhov — lexuslfa477@gmail.com
Alexei V. Stepanov — alexeistep45@mail.ru

ХРОНИКА

17 мая 2023 г. на заседании секции теоретической механики им. проф. Н. Н. Поляхова в Доме ученых им. М. Горького (Санкт-Петербург) в рамках мини-симпозиума по актуальным проблемам механики робототехнических систем заслушано два доклада.

1. Доклад кандидата физ.-мат. наук, доцента, вед. науч. сотр. М. З. Досаева, доктора физ.-мат. наук, профессора В. А. Самсонова и мл. науч. сотр. М. А. Гарбуза (НИИ механики МГУ имени М. В. Ломоносова) «Об отрыве вибрационного робота от шероховатой поверхности».

Краткое содержание доклада:

В докладе рассмотрено плоскопараллельное движение вибрационного робота, состоящего из корпуса, однородного маховика и дебаланса. Построена математическая модель системы. Показана принципиальная возможность управления дебалансом, результатом которого является подскок робота и его смещение в продольном направлении. Движение дебаланса разбито на несколько этапов, включая этап разгона, этап прыжка, замедление, снижение и приземление. Описаны условия полной остановки корпуса после совершения прыжка. Проанализировано смещение корпуса из начального положения.

2. Доклад кандидата физ.-мат. наук, вед. науч. сотр. Ю. Д. Селюцкого (НИИ механики МГУ имени М. В. Ломоносова) «О движении трехопорного робота с эксцентриком по плоскости в условиях анизотропного трения».

Краткое содержание доклада:

В докладе рассматривается динамика трехопорного робота, перемещающегося по шероховатой горизонтальной плоскости. Предполагается, что между плоскостью и опорами действует сила сухого трения, причем в двух опорах трение является изотропным, а в третьей — анизотропным. На корпусе робота установлен эксцентрик. Предложен закон управления угловым ускорением эксцентрика, обеспечивающий направленное движение робота. Исследуется влияние параметров системы (в частности, коэффициентов трения) и коэффициентов, входящих в закон управления, на характеристики движения.