УДК 519.2
MSC 62G30

Вестник СПбГУ. Математика. Механика. Астрономия. 2017. Т. 4 (62). Вып. 3

СРАВНЕНИЕ ЧИСЕЛ РЕКОРДОВ В ПОСЛЕДОВАТЕЛЬНОСТЯХ ДИСКРЕТНЫХ И НЕПРЕРЫВНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН*

В. Б. Невзоров

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

В ряде областей человеческой деятельности фиксируются различные рекордные достижения. Часто при этом происходит некоторая дискретизация с различной точностью (до секунды, до метра, до тысячи индивидуумов) получаемых результатов наблюдений. На примере экспоненциальных и геометрических распределений показано, как такого рода переход от непрерывных к дискретным вероятностным распределениям меняет число рекордных значений в соответствующих последовательностях случайных величин. Библиогр. 4 назв.

Ключевые слова: экспоненциальные распределения, геометрические распределения, рекордные моменты, рекордные величины.

1. Введение. В теории вероятностей и в ее многочисленных приложениях важное место отводится экспоненциально и геометрически распределенным случайным величинам. Во многом этот факт связан с присущим этим величинам так называемому свойству отсутствия последействия, которое позволяет в ряде случаев существенно упростить получение различных теорем. Многие важные результаты, связанные с этими распределениями, представлены, в частности, в теории порядковых статистик и в теории рекордов. Ниже мы обсудим ряд соотношений, связывающих рекордные величины для этих двух родственных (дискретных геометрических и непрерывных экспоненциальных) семейств случайных величин, обладающих указанным выше свойством.

2. Определения и обозначения. Рассмотрим последовательность независимых одинаково распределенных (н. о. р.) случайных величин (с. в.) $X_1, X_2, \ldots$. Определим для этой последовательности *рекордные моменты* $L(n)$, $n = 1, 2, \ldots$, и *рекордные величины* $X(1) < X(2) < \ldots$ следующим образом:

$$\begin{aligned} &L(1) = 1, \quad X(1) = X_1, \quad L(n+1) = \min\{j : X_j > X_{L(n)}\}, \quad n = 1, 2, \ldots, \\ &X(n) = X_{L(n)}, \quad n = 1, 2, \ldots \end{aligned} \tag{1}$$

Аналогичным образом можно определить для исходной последовательности $\{X\}$ слабые рекордные моменты $l(n)$ и слабые рекордные величины $\tilde{X}(n)$:

$$\begin{aligned} &l(1) = 1, \quad \tilde{X}(1) = X_1, \quad l(n+1) = \min\{j > l(n) : X_j \geq X_{l(n)}\}, \quad n = 1, 2, \ldots, \\ &\tilde{X}(n) = X_{l(n)}, \quad n = 1, 2, \ldots \end{aligned} \tag{2}$$

Слабые рекорды отличаются от введенных в (1) рекордов лишь тем, что в этом случае повторение (не обязательно превышение!) предыдущего рекордного значения также засчитывается как новое рекордное достижение. Имеет смысл рассматривать

*Работа выполнена при финансовой поддержке РФФИ (грант № 11-01-00769-а).

DOI: 10.21638/11701/spbu01.2017.308

слабые рекорды только для дискретных случайных величин, так как в случае непрерывных распределений совпадение каких-то двух с. в. $X_1, X_2, \ldots$ в исходной последовательности имеет нулевую вероятность.

Если быть более точными, то следует указать, что в (1) и (2) приведены определения *верхних рекордных моментов* $L(n)$, *верхних рекордных величин* $X(n)$, *слабых верхних рекордных моментов* $l(n)$ и *слабых верхних рекордных величин* $\tilde{X}(n)$. Дело в том, что аналогичным образом определяются и соответствующие нижние рекордные моменты и нижние рекордные величины. Для этого в соотношениях (1) и (2) знаки неравенств $>$ и $\geq$ достаточно заменить, соответственно, знаками $<$ и $\leq$. Результаты, полученные для верхних рекордов в последовательностях $X_1, X_2, \ldots$, практически сразу переносятся на нижние рекорды для последовательностей $W_1 = -X_1$, $W_2 = -X_2, \ldots$ Поэтому, как правило, если речь не идет о какой-то специальной ситуации, когда интерес представляют исключительно *нижние рекордные величины,* ограничиваются исследованием различных схем, связанных с *верхними рекордными моментами* и *верхними рекордными величинами,* которые и называют, как правило, просто рекордными моментами и рекордными величинами.

Ниже мы будем иметь дело с последовательностями н. о. р. случайных величин $\xi_1, \xi_2, \ldots$ и $Z_1, Z_2, \ldots$, имеющих стандартное $E(1)$-экспоненциальное распределение с функцией распределения (ф. р.) $F(x) = \max\{0, 1 - \exp(-x)\}$, $-\infty < x < +\infty$, и с последовательностями $Y_1, Y_2, \ldots$ и $V_1, V_2, \ldots$ независимых величин, имеющих геометрическое $\mathrm{Geom}(p)$-распределение, для которых

$$P\{Y_k = m\} = P\{V_k = m\} = (1-p)p^m, \quad m = 0, 1, 2, \ldots, 0 < p < 1.$$

В дальнейшем параметр p рассматриваемых геометрически распределенных случайных величин будет уточнен.

Наряду с последовательностями с. в. $Z_1, Z_2, \ldots$ и $Y_1, Y_2, \ldots$ будем рассматривать соответствующие им последовательности рекордных величин $Z(1) = Z_1 < Z(2) < \ldots$ и $Y(1) = Y_1 < Y(2) < \ldots$, а также будем иметь дело со слабыми рекордными величинами $\tilde{Y}(1) = Y_1 \leq \tilde{Y}(2) \leq \ldots$, построенными по последовательностям $\{Y\}$.

К классическим результатам теории рекордов (см., например, работы [1–4]) относятся следующие представления последовательностей заведомо зависимых рассматриваемых нами рекордных величин $Z(n)$, $Y(n)$ и $\tilde{Y}(n)$ в виде сумм независимых случайных слагаемых:

$$\{Z(1), Z(2), \ldots, Z(n)\} \stackrel{d}{=} \{\xi_1, \xi_1 + \xi_2, \ldots, \xi_1 + \xi_2 + \ldots + \xi_n\}, \quad n = 1, 2, \ldots; \tag{3}$$

$$\{Y(1), Y(2), \ldots, Y(n)\} \stackrel{d}{=} \{V_1, V_1 + V_2 + 1, \ldots, V_1 + V_2 + \ldots + V_n + n - 1\}, \quad n = 1, 2, \ldots; \tag{4}$$

$$\{\tilde{Y}(1), \tilde{Y}(2), \ldots \tilde{Y}(n)\} \stackrel{d}{=} \{V_1, V_1 + V_2, \ldots V_1 + V_2 + \ldots + V_n\}, \quad n = 1, 2, \ldots \tag{5}$$

В соотношениях (3)–(5) имеем дело с равенствами по распределению соответствующих векторов. Учитывая эти соотношения, получаем, что математические ожидания приведенных здесь рекордных величин задаются равенствами

$$EZ(n) = E(\xi_1 + \xi_2 + \ldots + \xi_n) = n, \quad n = 1, 2, \ldots; \tag{6}$$

$$EY(n) = E(V_1 + V_2 + \ldots + V_n + n - 1) = n/(1-p) - 1, \quad n = 1, 2, \ldots, \tag{7}$$

$$E\tilde{Y}(n) = E(V_1 + V_2 + \ldots + V_n) = np/(1-p), \quad n = 1, 2, \ldots \tag{8}$$

Отметим также, что из соотношений (3)–(5) получаем при $n \to \infty$ сходимость к стандартному $N(0,1)$-нормальному закону с функцией распределения

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{x} \exp(-t^2/2)dt$$

следующих случайных величин:

$$\frac{Z(n)-n}{\sqrt{n}}, \quad \left(Y(n) - \frac{n}{1-p}\right) \Big/ \left(\frac{np}{1-p}\right)^{1/2}, \quad \left(\tilde{Y}(n) - \frac{np}{1-p}\right) \Big/ \left(\frac{np}{1-p}\right)^{1/2}.$$

Ниже будут рассматриваться, наряду с экспоненциально распределенными с. в. Z_1, Z_2, ..., и величины, представляющие их целые части. Отметим несколько фактов для целой $Y = [Z]$ и дробной $W = Z - [Z]$ частей с. в. Z, имеющей $E(1)$-экспоненциальное распределение. Нетрудно убедиться, что

$$P\{Y = m\} = P\{m \leq Z < m+1\} = e^{-m}\left(1 - \frac{1}{e}\right), \quad m = 0, 1, 2, \ldots, \tag{9}$$

т. е. целая часть Z имеет геометрическое $\mathrm{Geom}(1/e)$-распределение. Дробная часть Z (величина W) сосредоточена на интервале $[0,1)$ и имеет ф. р.

$$F_W(x) = (1 - e^{-x})/(1 - 1/e), \quad 0 \leq x < 1, \tag{10}$$

т. е. $P\{W < x\} = P\{Z < x | Z < 1\}$. Нетрудно также убедиться, что с. в. Y и W независимы, поскольку

$$P\{Y = m, W < x\} = P\{m \leq Z < m + x\} = e^{-m} - e^{-(m+x)} = \\ = P\{Y = m\}P\{W < x\}, \quad m = 0, 1, 2, \ldots, \quad 0 \leq x < 1. \tag{11}$$

Таким образом, получили, что с. в. Z представима в виде суммы $Z = Y + W$ независимых случайных слагаемых, распределения которых определяются соотношениями (10) и (11).

3. Рекордные достижения в различных областях человеческой деятельности. Всевозможные рекорды трактуются населением нашей планеты как личные достижения, как результаты, свидетельствующие о прогрессе человеческой цивилизации, как стимул совершенствоваться в своей деятельности, чтобы выйти на новый рекордный уровень. Поэтому вполне понятен интерес к различным сообщениям о преодолении тех или иных рекордных рубежей. Многочисленные статьи в газетах, сообщения в Интернете, телепередачи немедленно дают знать о рекордах (даже, порой, экзотических), установленных в родной стране, в Европе, в мире или на космической орбите. Одной из самых популярных во всем мире является переводимая на многочисленные языки и регулярно (начиная с 1955 года) переиздаваемая книга, носящая названия «Книга рекордов Гиннесса» или «Гиннесс. Мировые рекорды».

Рассмотрим некоторые статистические проблемы, связанные с фиксацией рекордных результатов. Речь идет о тех ситуациях, когда могут расходиться предсказания о возможных будущих рекордных значениях, основанные на теоретических расчетах, и реальные сообщения о новых рекордных достижениях в том или ином виде человеческой деятельности. На примере упомянутых выше экспоненциальных и геометрических вероятностных распределений проанализируем, как процесс дискретизации исходных величин влияет на число рекордов в рассматриваемых последовательностях случайных величин. Вначале приведем взятые из Интернета некоторые рекордные (на момент начала 2017 года) достижения, фиксируемые жителями нашей планеты.

1. Уровни Невы выше ординара (приводимые в различных источниках с точностью до сантиметра) всегда волновали не только реальных жителей города на Неве, но и некоторых литературных героев. Ее последний рекордный уровень, составивший 421 см выше ординара, был зафиксирован 19 ноября 1824 года. Он заметно превзошел предыдущее рекордное значение, равное 310 см, которое было отмечено в 1777 году.
2. С июня 2013 года рекордное достижение по продолжительности жизни мужчин на Земле принадлежит японцу Кимуре. Он прожил 116 лет и 54 дня. Предыдущий рекорд, составивший 115 лет и 252 дня держался с 1998 года. В этой ситуации не было возможности, да и это не требовалось, уточнять рекордную продолжительность жизни с точностью, скажем, до часа.
3. Предыдущий рекордный результат из отмеченных в различных записях за последнее столетие, полученный при измерении роста человека, был зафиксирован в США в 1921 году и равнялся 248 см. Этот рекорд был превышен на 24 сантиметра другим американским великаном в 1940 году. Видим, что и в этом случае не было необходимости измерять рост с точностью до миллиметра.
4. Самая низкая температура на Земле (с момента, когда началась ее запись) была зафиксирована в 1983 году на станции Восток в Антарктиде. Она составила $-89,2$ градуса. В Европе как рекордно жаркая за все времена измерений была отмечена в 1881 году температура воздуха $+50,0$ в испанской Севилье. В этих случаях уже была возможность фиксировать измерения с точностью до одной десятой доли градуса.

Видим, что в приведенных примерах появление нового рекордного достижения представляет собой весьма редкое событие. Округление наблюдаемых величин до одного дня, одного сантиметра, десятой доли градуса совсем не отразилось на числе рекордных значений. Совсем другая ситуация наблюдается, если перейти к спортивным результатам. Здесь громадное число участников многочисленных соревнований приводит к большой плотности результатов. В некоторых видах легкой атлетики решающей бывает разница в несколько сантиметров или сотых долей секунды. В качестве примеров можно взять результаты спортсменов на спринтерских дистанциях.

Рассмотрим, например, таблицу мировых рекордов на стометровой дистанции у мужчин. До 1975 года использовался ручной хронометраж. Секундомеры позволяли фиксировать результаты бегунов с точностью до десятых долей секунды. В этой ситуации результат 10,0 секунды с 1960 по 1968 год был зафиксирован как рекордный 14 раз. Затем очередные 8 лет привели к установлению и вслед за этим 15-кратному повторению рекордного значения 9,9 секунды, что дало миру большое число сорекордсменов на этой дистанции. Переход с 1 января 1975 года на электронный хронометраж, позволивший уже с точностью до сотых долей секунды фиксировать результаты спортсменов, снизил число «слабых» рекордсменов, увеличив число обладателей классических рекордов. Среди новых рекордных значений были, в частности, последовательно зафиксированы результаты 9,86, 9,85, 9,84, 9,79, 9,77 (4 раза!), 9,74, 9,72, 9,69. Наконец, в 2009 году Усэйн Болт, представлявший Ямайку, немного оторвался от предыдущего результата и показал остающийся рекордным уже 8 лет результат 9,58 секунды. Видим, что даже и такой более точный хронометраж привел к появлению слабых рекордных значений. Уже даже сделаны (пока еще неофициальные) попытки фиксировать в этом виде легкой атлетики результаты, округленные до тысячных долей секунды. В частности, приводится и более точное значение результата Болта — 9,578 секунды. Интересно, что в некоторых видах спорта уже официально перешли на точность, составляющую тысячные доли секунды. Например, в велогонках на треке рекордный результат Франсуа Первиса на дистанции 200 метров с хода равен 9,347 секунды.

Во всех приведенных выше примерах, по сути дела, вместо наблюдаемых значений абсолютно непрерывных случайных величин рассматриваются их дискретные аналоги, полученные использованием того или иного варианта округления истинных величин. Ниже на примере экспоненциальных и полученных в результате их дискретизации геометрических распределений будут проанализированы возможные взаимоотношения рекордных величин, возникающие при переходе от непрерывных распределений к дискретным.

4. Соотношения между рекордами при переходе от экспоненциальных к геометрическим случайным величинам. Рассмотрим в качестве исходной последовательность н. о. р. случайных величин Z_1, Z_2, ..., имеющих стандартное $E(1)$-экспоненциальное распределение с ф. р. $F(x) = \max\{0, 1 - \exp(-x)\}$, $-\infty < x < \infty$. Переход от этих величин к соответствующим дискретным аналогам можно провести различными способами.

Начнем с ситуации, когда вместо исходных величин рассматриваются их целые части $Y_1 = [Z_1]$, $Y_2 = [Z_2]$, ... Эти случайные величины принимают целые значения $0, 1, 2, \ldots$ и имеют геометрическое $\mathrm{Geom}(p)$-распределение с параметром $p = 1/e$. Наряду с рекордными в последовательности Z_1, Z_2, ... случайными величинами $Z(1) = Z_1 < Z(2) < \ldots$ будем рассматривать также рекорды $Y(1) = Y_1 < Y(2) < \ldots$ и слабые рекордные величины $\tilde{Y}(1) = Y_1 \le \tilde{Y}(2) \le \ldots$, порожденные последовательностью целочисленных с. в. $Y_1, Y_2, \ldots$

Простые рассуждения позволяют убедиться, что между тремя приведенными здесь наборами рекордных величин справедливы следующие соотношения:

$$\tilde{Y}(n) \le Z(n) \le Y(n), \quad n = 1, 2, \ldots,$$

и

$$E\tilde{Y}(n) \le EZ(n) \le EY(n), \quad n = 1, 2, \ldots$$

Зная из равенства (6), что $EZ(n) = n$, интересно сравнить это значение с математическими ожиданиями рекордных величин $\tilde{Y}(n)$ и $Y(n)$. Для этого нужно воспользоваться соотношениями (7) и (8) при $p = 1/e$. Получаем

$$EY(n) = \frac{ne}{e-1} - 1, \quad n = 1, 2, \ldots,$$

и

$$E\tilde{Y}(n) = \frac{n}{e-1}, \quad n = 1, 2, \ldots,$$

т. е. при больших значениях n величина $EY(n)$ примерно в 1,72 раза больше математического ожидания $EZ(n)$, а значение $E\tilde{Y}(n)$ примерно в 1,58 раза меньше среднего значения рекордной величины $Z(n)$.

Если рассмотреть в этой ситуации с. в. $R_1(n)$, представляющие собой числа рекордов в наборах Z_1, Z_2, …, Z_n, а также с. в. $R_2(n)$ и $R_3(n)$, совпадающие с количеством рекордов и слабых рекордов среди величин $Y_1, Y_2, \ldots, Y_n$, $n = 1, 2, \ldots$, то можно показать, что при $n \to \infty$ распределения отношений $R_2(n)/R_1(n)$ и $R_3(n)/R_1(n)$ сходятся, соответственно, к вырожденным в точках $1/(e-1)$ и $e/(e-1)$ распределениям.

Замечание. Более значительные взаимоотношения между рассматриваемыми рекордами можно наблюдать, если в качестве исходной взять последовательность н. о. р. с. в. Z_1, Z_2, …, имеющих произвольное $E(\lambda)$-распределение с ф. р. $F(x) = 1 - \exp(-x/\lambda)$, $x \geq 0$, в ситуации, когда параметр $\lambda > 0$ принимает большие значения. В этом случае соответствующие распределения отношений $R_2(n)/R_1(n)$ и $R_3(n)/R_1(n)$ будут сходиться к вырожденным распределениям в точках $1/(e^{1/\lambda} - 1)$ и $e^{1/\lambda}/(e^{1/\lambda} - 1)$. Чем больше значение параметра λ, тем существеннее различия между величинами $R_1(n)$, $R_2(n)$ и $R_3(n)$. Поскольку $(e^{1/\lambda} - 1) \sim 1/\lambda$ при стремящихся к бесконечности значениях λ, то видим, что за счет выбора уровня дискретизации можно получить сколь угодно большие значения отношений $R_2(n)/R_1(n)$ и $R_3(n)/R1(n)$. Получаем, что рассмотренный процесс дискретизации случайных величин может привести к существенным изменениям числа рекордов, которые могли бы быть зафиксированными в исходной последовательности величин.

Рассмотрим еще одну ситуацию, когда вместо с. в. Z_1 Z_2, … будем брать не их целые части, а их округления до ближайших к ним целым значениям. Для того чтобы иметь дело с более простыми выкладками, вместо этих величин будем рассматривать в качестве исходной последовательность с. в. $T_n = Z_n - 1/2, \ldots, n = 1, 2, \ldots$, что не изменит соответствующего числа рекордов, фиксируемых среди с. в. Z_1, Z_2, … Если обозначим через $T(1) = T_1 < T(2) < \ldots$ соответствующие рекордные величины, для которых справедливы равенства $T(n) = Z(n) - 1/2$, $n = 1, 2, \ldots$, то из соотношения (3) получаем

$$\{T(1), T(2), \ldots, T(n)\} \stackrel{d}{=} \left\{\xi_1 - \frac{1}{2}, \xi_1 + \xi_2 - \frac{1}{2}, \ldots, \xi_1 + \xi_2 + \ldots + \xi_n - \frac{1}{2}\right\}, \quad n = 1, 2, \ldots$$

В этой ситуации $ET(n) = n - 1/2$, $n = 1, 2, \ldots$

Обозначим через X_n случайную величину, полученную округлением T_n до ближайшего целого числа. Пусть $X(1) = X_1 < X(2) < \ldots$ представляют рекордные значения в последовательности с. в. $X_1, X_2, \ldots$ Видим, что

$$P\{X_n = m\} = P\left\{m - \frac{1}{2} \leq T_n < m + \frac{1}{2}\right\} =$$
$$= P\{m \leq Z_n < m + 1\} = e^{-m}\left(1 - \frac{1}{e}\right), \quad m = 0, 1, 2, \ldots,$$

т. е. при каждом $n = 1, 2, \ldots$ с. в. X_n имеют такое же геометрическое $\mathrm{Geom}(1/e)$-распределение, что и с. в. $Y_n = [Z_n]$, представленные выше. Получаем, вспоминая, в частности, равенство (4), что справедливы следующие соотношения:

$$\{X(1), X(2), \ldots, X(n)\} \stackrel{d}{=} \{X_1, X_1 + X_2 + 1, \ldots, X_1 + X_2 + \ldots + X_n + n - 1\},$$

$$EX(n) = \frac{ne}{e-1} - 1, \quad n = 1, 2, \ldots$$

Отметим также, что при $n \to \infty$ последовательность с. в.

$$\left(X(n) - \frac{ne}{e-1}\right) \Big/ \left(\frac{n}{e-1}\right)^{1/2}$$

сходится по распределению к стандартному $N(0,1)$-нормальному закону.

Таким образом, и такой способ получения дискретных величин, близких к исходным, имеющим непрерывные функции распределения, существенным образом сказывается на числе рекордных значений.

Литература

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

2. *Chandler K. N.* The distribution and frequency of record values // J. R. Stat. Soc., Ser. B. 1952. Vol. 14. P. 220–228.

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

4. *Vervaat W.* Limit theorems for records from discrete distributions // Stochast. Process. Appl. 1973. Vol. 1, N 2. P. 317–334.

Статья поступила в редакцию 6 февраля 2017 г.; рекомендована в печать 30 марта 2017 г.

Сведения об авторе

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

COMPARISON OF NUMBERS OF RECORDS IN THE SEQUENCES OF DISCRETE AND CONTINUOUS RANDOM VARIABLES

Valery B. Nevzorov

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

Different record achievements are fixed in many domains of human activities. Very often this process happens with some rate of discretization (up to seconds, up to meters or up to thousands of individuals) of observed results. Using the examples of exponential and geometrical distributions, it is shown how such type of transition from continuous to discrete distributions can vary the numbers of record values in corresponding sequences of random variables. Refs 4.

Keywords: exponential distributions, geometrical distributions, record times, record values.

References

1. Arnold B. C., Balakrishnan N., Nagaraja H. N., *Records* (John Wiley & Sons, New York, 1998, 312 p.).

2. Chandler K. N., “The distribution and frequency of record values”, *J. R. Stat. Soc., Ser. B* **14**, 220–228 (1952).

3. Nevzorov V. B., *Records. Mathematical Theory.* In Ser. *Translations of Mathematical Monographs* (American Math. Society, **194**, 2001).

4. Vervaat W., “Limit theorems for records from discrete distributions”, *Stochast. Process. Appl.* **1**(2), 317–334 (1973).

Для цитирования: Невзоров В. Б. Сравнение числа рекордов в последовательностях дискретных и непрерывных случайных величин // Вестник СПбГУ. Математика. Механика. Астрономия. 2017. Т. 4 (62). Вып. 3. С. 459–465. DOI: 10.21638/11701/spbu01.2017.308

For citation: Nevzorov V. B. Comparison of numbers of records in the sequences of discrete and continuous random variables. *Vestnik SPbSU. Mathematics. Mechanics. Astronomy*, 2017, vol. 4 (62), issue 3, pp. 459–465. DOI: 10.21638/11701/spbu01.2017.308