УДК 519.2 Вестник СПбГУ. Математика. Механика. Астрономия. 2021. Т. 8 (66). Вып. 1
MSC 62G32

Об обобщениях задачи оптимального выбора*

И. В. Бельков

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

Для цитирования: *Бельков И. В.* Об обобщениях задачи оптимального выбора // Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия. 2021. Т. 8 (66). Вып. 1. С. 29–36. https://doi.org/10.21638/spbu01.2021.103

В статье рассматриваются обобщения задачи оптимального выбора. Имеется последовательность из n независимых случайных величин, одинаково распределенных на отрезке $[0, 1]$. Последовательно получая наблюдаемые значения этих величин, нужно в какой-то момент остановиться на одной из них, приняв ее как начальную для отсчета верхних или нижних рекордных величин. В оптимизационных задачах требуется сделать правильный выбор начальной точки отсчета рекордов, чтобы максимизировать математическое ожидание суммы значений или количества верхних, нижних или тех и других рекордных величин, полученных в результате такой процедуры. Приводятся обзор результатов, посвященных равномерному распределению случайных величин, и новые результаты, касающиеся экспоненциального распределения случайных величин.

Ключевые слова: рекордные величины, выборочные размахи, рекордные размахи, экспоненциальное распределение, распределение Лапласа, геометрическое распределение.

1. Классическая задача оптимального выбора (другие варианты названия — задача о разборчивой невесте, проблема секретаря) была рассмотрена М. Гарднером в работе [1].

Сущность задачи заключается в следующем. Невесте (или работодателю) требуется выбрать жениха (или секретаря) среди заранее известного числа n претендентов. Невеста общается с претендентами в случайном порядке (с каждым из них не более одного раза). О каждом претенденте известно, лучше или хуже он всех предыдущих по какому-либо фактору, выраженному количественно (например, по зарплате). В результате общения с претендентом невеста должна либо принять, либо отклонить его предложение. Какой стратегии должна придерживаться невеста, чтобы максимизировать вероятность выбора наилучшего жениха?

Математически эта задача формулируется таким образом. Пусть существует последовательность случайных величин $X_1, X_2, ..., X_n$. Выделим из них случайное число $M = M(n)$ верхних рекордных величин:

$$X(1) = X_1 = X(2) < ... < X(M) = \max\{X_1, X_2, ..., X_n\}.$$

Последовательно получаем наблюдаемые значения $x_1, x_2, ...$ этих случайных величин. Получив очередное наблюдение x_k, нужно решить, останавливаться ли на этом наблюдении или продолжить процесс выбора в надежде выйти позже на оптимальное значение наблюдаемой величины.

*Работа выполнена при финансовой поддержке РФФИ (грант № 18-01-00393).

Нужно угадать момент появления последнего рекорда в последовательности $X_1, X_2, ..., X_n$.

Верная стратегия была найдена Е. Б. Дынкиным в работе [2]. В основу заложен принцип динамического программирования. Нужно при фиксированном числе n элементов выборки пропустить какое-то число r наблюдений $x_1, x_2, \dots, x_r$, а затем выбрать первое значение x_s, $s > r$, которое превышает все предыдущие. Существуют таблицы, позволяющие при любом значении $n = 2, 3, \dots$ находить величины $r = r(n)$ и вероятности $p(n)$, позволяющие угадать момент появления последнего рекорда в наборе $X_1, X_2, \dots, X_n$. В частности, при $n \to \infty$ получаем $r(n) \sim \frac{n}{e}$ и $p(n) \to \frac{1}{e} = 0.367879...$

Существуют также различные задачи, связанные с проблемой оптимального выбора. В. Б. Невзоров и С. А. Товмасян в статье [3] исследовали вопрос об увеличении числа рекордов в последовательностях независимых одинаково распределенных случайных величин в ситуации, когда в качестве начала отсчета можно выбрать какую-то из величин X_k (не обязательно X_1). В статье [4] В. Б. Невзорова и И. В. Белькова также была рассмотрена задача об увеличении числа рекордов в последовательности случайных величин $X_1, X_2, ..., X_n$ при применении процедуры, использованной в задаче о разборчивой невесте. В статье [5] тех же авторов была рассмотрена задача о максимизации математического ожидания суммы верхних рекордных величин в случае, когда исходные независимые величины $U_1, U_2, ..., U_n$ имеют равномерное распределение на интервале [0, 1].

2. Обозначим через $T(n)$ математическое ожидание суммы рекордных величин в наборе $U_1, U_2, ..., U_n$, которое равно

$$T(n) = \frac{1}{2} + ... + \frac{1}{n} + \frac{1}{n+1}.$$

Через $T_n(x)$ будем обозначать математическое ожидание в наборе $x, U_2, ..., U_n$:

$$T_n(x) = \frac{1}{2} + ... + \frac{1}{n} + x - \frac{x^2}{2} - \frac{x^3}{3} - ... - \frac{x^n}{n}.$$

Зададимся целью максимально увеличить математическое ожидание суммы рекордных значений среди величин $U_1, U_2, ..., U_n$ за счет правильного выбора в этом наборе начальной точки отсчета рекордов. Если начальная точка U_1 в этом наборе принимает значение, близкое к единице, то это значение вносит существенный вклад в сумму $T(n)$, но остальные величины имеют значительно меньше шансов стать рекордными и внести свой вклад в сумму $T(n)$. Остается надеяться, что если не принимать в расчет эту случайную величину, а перейти сразу к набору $U_2, U_3, ..., U_n$, то можно увеличить математическое ожидание суммы рекордных величин. Каким должно быть значение наблюдаемой случайной величины U_1, чтобы определить, учитывать ли его как первое рекордное или перейти к набору $U_2, U_3, ..., U_n$? Если $n = 1$ или $n = 2$, то никакие переходы не могут увеличить $T(n)$.

Пусть x_n — значения корней уравнения $T(n-1) = T_n(x)$; V_n — максимизированное математическое ожидание суммы рекордных величин в случае, если отсчет начинать не с первой, а с другой величины; $d(n) = V_n - T(n)$.

Тогда следует начать со сравнения при каждом фиксированном $n = 2, 3, ...$ и любом $x \in (0, 1)$ величины $T_n(x)$ и математического ожидания суммы рекордов в наборе $U_2, ..., U_n$, которое совпадает с $T(n-1)$.

Из $T(n-1) = T_n(x)$ следует уравнение

$$\frac{x}{2} + \frac{x^2}{3} + ... + \frac{x^{n-1}}{n} = 1. \qquad (1)$$

В результате получаем следующую таблицу значений x_n — корней уравнения (1) при различных n:

n	x_n
4	0.9554
5	0.8852
6	0.8509
7	0.8318
8	0.8204
9	0.8131
10	0.8084

Таким образом, если значение $x > x_n$, то отсчет рекордов начинается с U_1. Если $x < x_n$, то следует пропустить U_1 и перейти к следующей величине. Здесь x — значение величины U_1.

В статье [5] приведены также таблицы значений величин $V_n, T(n)$ и $d(n)$.

3. Теперь рассмотрим аналогичную ситуацию для нижних рекордов [4].

Нижние рекорды определяются следующим образом:

$$X(1) = X_1 > X(2) > \cdots > X(M) = \min\{X_1, X_2, ..., X_n\}.$$

В этом случае будем через $T(n)$ обозначать математическое ожидание суммы нижних рекордных значений в наборе $U_1, U_2, ..., U_n$, а через $T_n(x)$ — математическое ожидание суммы нижних рекордных значений в наборе $x, U_2, ..., U_n$:

$$T(n) = 1 - \frac{1}{n+1}, \quad n = 1, 2, ...; \quad T_n(x) = 2x - \frac{1}{n} + \frac{(1-x)^n}{n}.$$

Тогда из равенства $T(n-1) = T_n(x)$ следует уравнение

$$2x + \frac{(1-x)^n}{n} = 1, \quad n = 2, 3, ... \qquad (2)$$

Приводим таблицу корней уравнения (2) на интервале (0, 1):

n	x_n
2	0.414214
3	0.476024
4	0.496153
5	0.496773
6	0.498677
7	0.499438
8	0.499755
9	0.499891
10	0.499990

В случае $x \geq x_n$ следует брать величину U_1, а в случае $x < x_n$ нужно отбросить значение U_1 и перейти к следующей величине. Здесь через x обозначено значение величины U_1.

4. Задача о максимизации суммы всех (верхних и нижних) рекордных значений была рассмотрена в статье И. В. Белькова [6].

В этом случае через $T(n)$ будет обозначаться математическое ожидание суммы всех (верхних и нижних) рекордных значений в наборе $U_1, U_2, ..., U_n$, а через $T_n(x)$ — математическое ожидание суммы всех нижних рекордных значений в наборе $x, U_2, ..., U_n$.

Уравнение $T(n-1) = T_n(x)$ имеет следующий вид:

$$2x + \frac{(1-x)^n}{n} - \frac{x^2}{2} - \frac{x^3}{3} - ... - \frac{x^n}{n} = \frac{1}{2}. \tag{3}$$

Таблица корней уравнения (3) такова:

n	$(x_n)_1$	$(x_n)_2$
2	0	
3	0.15693	
4	0.21663	
5	0.24273	
6	0.25560	
7	0.26252	0.97522
8	0.26647	0.94983
9	0.26882	0.93301
10	0.27026	0.92135
20	0.27284	0.89105

Отметим, что ситуации оптимального выбора здесь отличаются в случаях, когда $n = 3, 4, 5, 6$ и когда $n \geq 7$.

Если $n = 3, 4, 5, 6$, то первое наблюдаемое значение x, если оно больше x_n, выбираем в качестве начала отсчета; если же оно меньше x_n, то его отбрасываем. Если $n \geq 7$, то в случае $(x_n)_1 \leq x \leq (x_n)_2$ величину x берем в качестве первого рекорда; иначе ее отбрасываем и переходим к следующей случайной величине.

Задача об увеличении математического ожидания суммарного числа верхних и нижних рекордов [4] приводит к уравнению

$$2 = \left(x + \frac{x^2}{2} + \frac{x^3}{3} + ... + \frac{x^{n-1}}{n-1}\right) + \left((1-x) + \frac{(1-x)^2}{2} + \frac{(1-x)^3}{3} + ... + \frac{(1-x)^{n-1}}{n-1}\right). \tag{4}$$

Отметим, что при $n = 1, 2, 3, 4$ корней уравнения на $[0,\ 1]$ нет. Для $n \geq 5$ получаем следующие значения критических уровней:

n	$(x_n)_1$	$(x_n)_2$
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

Если испытаний не более четырех, то следует начинать отсчет со случайной величины U_1. Если испытаний пять или более, то в случае $(x_n)_1 < U_1 < (x_n)_2$ принимаем U_1 за начало отсчета, иначе переходим к U_2.

5. Попытаемся теперь рассмотреть случай, когда исходные величины имеют стандартное экспоненциальное распределение.

Пусть независимые случайные величины $X_1, X_2, ..., X_n$ имеют E(1)-экспоненциальное распределение с функцией распределения

$$F(x) = \begin{cases} 1 - e^{-x}, \ x \geq 0; \\ 0, x < 0 \end{cases}$$

и плотностью распределения

$$f(x) = \begin{cases} e^{-x}, x \geq 0; \\ 0, x < 0. \end{cases}$$

Зададимся целью максимально увеличить в этом наборе математическое ожидание суммы верхних рекордных значений.

Рассмотрим ситуацию, когда первая величина имеет значение $x > 0$.

Обозначим через $T(n)$ математическое ожидание суммы верхних рекордных значений в наборе $X_1, X_2, ..., X_n$, а через $T_n(x)$ — математическое ожидание суммы верхних рекордных значений в наборе $x, X_2, ..., X_n$.

Найдем формулы для значений величин $T(n)$ и $T_n(x)$.

Если величины имеют распределение с произвольной функцией распределения $F(x)$, то получаем для $T(n)$ и $T_n(x)$ следующие выражения:

$$T(n) = \sum_{k=1}^{n} xF^{k-1}(x)dF(x) = \int_{-\infty}^{\infty} x\frac{1 - F^n(x)}{1 - F(x)}dF(x);$$

$$T_n(x) = x + \sum_{k=1}^{n-1} \int_x^1 uF^{k-1}(u)dF(u).$$

Поставив сюда функцию и плотность экспоненциального распределения, получаем

$$T(n) = \int_0^{+\infty} x(1 - (1 - e^{-x})^n)dx = \int_0^{+\infty} x\left(1 - \sum_{k=0}^{n} C_n^k(-e^{-x})^k\right)dx =$$

$$= \int_0^{+\infty} \left(x - \sum_{k=0}^{n} C_n^k(-e^{-x})^k x\right)dx =$$

$$= \left(\sum_{k=1}^{n} C_k^n \frac{(-e^{-x})^k(kx+1)}{k^2}\right)\Bigg|_0^{+\infty} = -\sum_{k=1}^{n}(-1)^k\frac{C_n^k}{k^2};$$

$$T_n(x) = x + \sum_{k=1}^{n-1}\int_x^{+\infty} ue^{-u}\sum_{k=1}^{n-1}(1 - e^{-u})^{k-1}e^{-u}du =$$

$$= x + \int_x^{+\infty} ue^{-u}\sum_{k=1}^{n-1}(1 - e^{-u})^{k-1}e^{-u}du = x + \int_x^{+\infty} u(1 - (1 - e^{-u})^{n-1})du.$$

Получаем следующую таблицу значений $T(n)$:

n	$T(n)$
2	1.7500
3	2.3611
4	2.8819
5	3.3386
6	3.7469
7	4.1164
8	4.4571
9	4.7714
10	5.0643

Рассмотрим разности

$$R_n(x) = T(n-1) - T_n(x), \quad n = 2, 3, \ldots .$$

Если $U_1 = x$ и $R_n(x) \le 0$, то отсчет рекордов начинаем с величины U_1. Если же $R_n(x) > 0$, то следует отбросить U_1 и перейти к величине U_2. Затем определяем знак величины $R_{n-1}(x) = T(n-2) - T_{n-1}(x)$, при необходимости переходим к U_3 и т. д.

Рассмотрим уравнение

$$T(n-1) = T_n(x), \quad n = 2, 3, ...,$$

которое эквивалентно $R_n(x) = 0$.

Получаем

$$\int_0^{+\infty} u(1-(1-e^{-u})^{n-1})du = x + \int_x^{+\infty} u(1-(1-e^{-u})^{n-1})du.$$

Отсюда следует уравнение

$$\int_0^x u(1-(1-e^{-u})^{n-1})du = x,$$

которое также можно представить в виде

$$\sum_{k=1}^{n-1} C_{n-1}^k \frac{(-e^{-x})^k(kx+1)-1}{k^2} = x.$$

Данное уравнение имеет единственное неотрицательное решение, равное нулю. Следовательно, счет следует начинать с величины U_1 независимо от числа испытаний, т. е. в этом случае увеличить математическое ожидание суммы рекордов за счет переноса начала отсчета рекордов не удается.

Литература

1. Gardner M. Mathematical Games. A fifth collection of "brain-teasers". *Scientific American* **202** (2), 150–154 (1960).

2. Дынкин Е. Б. Оптимальный выбор момента остановки процесса. *Докл. АН СССР* **150** (2), 238–240 (1963).

3. Невзоров В. Б., Товмасян С. А. О максимальном значении среднего числа рекордов. *Вестник Санкт-Петербургского университета. Серия 1. Математика. Механика. Астрономия* **1 (59)**, вып. 2, 196–200 (2014).

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

5. Бельков И. В., Невзоров В. Б. Об одной задаче оптимального выбора рекордных величин. *Записки научн. сем. ПОМИ* **466**, 30–37 (2017).

6. Бельков И. В. О некоторых задачах оптимального выбора рекордных величин. *Научный журнал «Globus»*, вып. 11 (44), 46–49 (2019).

Статья поступила в редакцию 28 августа 2020 г.;
после доработки 15 сентября 2020 г.;
рекомендована в печать 17 сентября 2020 г.

Контактная информация:

Бельков Игорь Владимирович — аспирант; igor.belkov@gmail.com

On generalizations of the optimal choice problem*

I. V. Belkov

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

For citation: Belkov I. V. On generalizations of the optimal choice problem. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy*, 2021, vol. 8 (66), issue 1, pp. 29–36. https://doi.org/10.21638/spbu01.2021.103 (In Russian)

The article is dedicated to some generalizations of the classical optimal choice problem (the fastidious bride problem, the secretary problem). Let there be a sequence of n identically distributed random variables on the interval [0, 1]. Getting consistently observed values of these variables, we should stop at some moment on one of them, accepting it as the initial point for counting upper or lower record values. In the optimal choice problem and its generalizations, it is needed to make the correct choice of the initial point of counting records, in order to guess the place of the last record (the classical optimal choice problem) or to maximize the expected sum of upper and/or lower record values or the expected total number of upper and/or lower records, obtained by this procedure. A review of results on the uniform distribution of the random variables and some new results concerning the exponential distribution are presented.

Keywords: record moments, record values, sums of record values, mean number of records, uniform distribution, exponential distribution, optimal choice problem.

References

1. Gardner M. Mathematical Games. A fifth collection of “brain-teasers”. *Scientific American* **202** (2), 150–154 (1960).

2. Dynkin E. B. Optimal choice of the stopping time of a Markov process. *Dokl. Akad. Nauk SSSR* **150** (2), 238–240 (1963). (In Russian) [Engl. transl.: *Soviet Math. Dokl.* **4**, 627–629 (1963)].

3. Nevzorov V. B., Tovmasyan S. A. On the maximal value of the expectation of record numbers. *Vestnik of Saint Petersburg University. Series 1. Mathematics. Mechanics. Astronomy* **1 (59)**, iss. 2, 196–200 (2014). (In Russian) [Engl. transl.: *Vestnik St. Petersb. Univ. Math.* **47**, iss. 2, 64–67 (2014). https://doi.org/10.3103/S1063454114020046].

4. Belkov I. V., Nevzorov V. B. On the problem of the optimal choice of record values. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy* **5 (63)**, iss. 2, 179–188 (2018). https://doi.org/10.21638/11701/spbu01.2018.201 (In Russian) [Engl. transl.: *Vestnik St. Petersb. Univ. Math.* **51**, iss. 2, 107–113 (2018). https://doi.org/10.3103/S1063454118020024].

*The work is supported by Russian Foundation for Basic Research (grant no. 18-01-00393).

5. Bel'kov I. V., Nevzorov V. B. On one problem of the optimal choice of record values. *Zap. Nauchn. Sem. POMI* **466**, 30–37 (2017). (In Russian) [Engl. transl.: *J. Math. Sci.* **244**, 718–722 (2020). https://doi.org/10.1007/s10958-020-04644-0].

6. Belkov I. V. About some tasks of optimal selection of record values. *Scientific journal "Globus"*, iss. 11 (44), 46–49 (2019). (In Russian)

Received: August 28, 2020
Revised: September 15, 2020
Accepted: September 17, 2020

Author's information:

Igor V. Belkov — igor.belkov@gmail.com