Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Московский государственный университет имени М. В. Ломоносова»

На правах рукописи



Орлов Олег Павлович

Комбинаторно-вероятностные задачи для случайных выборок

Специальность 1.1.4 —
«Теория вероятностей и математическая статистика»

Автореферат
диссертации на соискание учёной степени
кандидата физико-математических наук

Москва — 2022

Работа выполнена на кафедре математической статистики и случайных процессов механико-математического факультета Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный университет имени М. В. Ломоносова».

Научный руководитель:
Зубков Андрей Михайлович - доктор физико-математических наук, заведующий отделом дискретной математики Федерального государственного бюджетного учреждения науки Математического института им. В.А. Стеклова Российской академии наук (специальность – 01.01.05)

Официальные оппоненты:
Топчий Валентин Алексеевич - доктор физико-математических наук, профессор, ведущий научный сотрудник Лаборатории комбинаторных и вычислительных методов алгебры и логики Омского филиала Федерального государственного бюджетного учреждения науки Института математики им. С.Л. Соболева Сибирского отделения Российской академии наук (специальность – 01.01.05)

Цитович Иван Иванович - доктор физико-математических наук, доцент, главный научный сотрудник Лаборатории № 1 им. М.С. Пинскера Федерального государственного бюджетного учреждения науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук (специальность – 05.13.17 теоретические основы информатики)

Ведущая организация:
Научно-исследовательский институт прикладных проблем математики и информатики Белорусского государственного университета

Защита состоится 22 декабря 2022 г. в 14 часов на заседании диссертационного совета 24.1.167.01 при Математическом институте им. В. А. Стеклова Российской академии наук по адресу: 119991, г. Москва, ул. Губкина, д. 8.

С диссертацией можно ознакомиться в библиотеке Математического института имени В. А. Стеклова РАН и на сайте http://www.mi-ras.ru/dis/ref22/orlov/dis.pdf

Автореферат разослан « » октября 2022 г.

Ученый секретарь
диссертационного совета
24.1.167.01,
доктор физико-математических наук В. А. Ватутин

Общая характеристика работы

Актуальность темы. Диссертация посвящена трем задачам комбинаторно-вероятностного типа. В первой главе диссертации изучаются межрекордные наполнения — новая характеристика в теории рекордов, вторая глава посвящена экстремальным расстояниям до ближайшего соседа, в третьей главе рассмотрена задача, связанная с классической схемой случайных размещений.

1) Математическая теория рекордов берет свое начало со статьи Чендлера 1952-го года [5]. Эта статья простимулировала появление работ [6—9], в которых уже рассматривались вопросы, связанные с применением рекордов для проверки некоторых статистических гипотез. Для этого в данной последовательности случайных величин выделяются рекорды, т. е. элементы, большие чем все предыдущие, моменты их появления и длины промежутков между этими моментами (межрекордные интервалы). Рекорды тесно связаны с порядковыми статистиками, особенно с крайними членами вариационного ряда. Теория экстремальных порядковых статистик достаточно широко изложена, например, в [10—12]. За 70 лет появилось большое число работ, посвященных рекордам. Несколько монографий опубликовано по теории рекордов, например, [13—15]. Обзорный характер носят статьи [16—18]. Следует выделить статьи Реньи [19], Таты [20] и Шоррока [21—23], которые способствовали интенсивному развитию математической теории рекордов.

Математическая теория рекордов используется для моделирования в спорте (см., например, [24; 25]), страховании (см., например, [26; 27]), гидрологии (см., например, [28, chapter 12]) и.т.д.

Введенные в диссертации величины — межрекордные наполнения — тесно связаны с рекордными моментами и межрекордными интервалами; межрекордные наполнения являются новыми характеристиками структуры случайных последовательностей.

2) Расстояния до ближайших соседей используются, например, при построении статистических критериев равномерности распределения (см., например, [29—32]), при оценке энтропии (см., например, [33; 34]), в алгоритмах классификации (см., например, [35]).

Случайные величины, связанные с расстояниями до ближайших соседей, изучались многими авторами (см., например, [36—39]). Главное отличие полученных в диссертации предельных теорем для максимального расстояния до ближайшего соседа от известных результатов заключается в том, что метрическое пространство, из которого взята выборка, может быть дискретным. В диссертации подробно рассмотрен случай двоичного куба с метрикой Хемминга, для которого получено предельное распределение максимального расстояния до ближайшего соседа в широком диапазоне изменения параметров.

3) "Классическая задача о дробинках" формулируется так: Пусть равновероятно и независимо друг от друга n частиц размещаются по N ячейкам. Какое распределение имеет случайная величина $\mu_0 = \mu_0(n, N)$, равная числу пустых ячеек?

Легко вычислить, что $\mathsf{E}\mu_0 = N(1 - 1/N)^n$. При $n, N \to \infty$ выделяют пять областей, связывающих n и N:

- левая область: $n^2/(2N) \to \lambda < \infty$, где λ — константа,
- левая промежуточная область: $n/N \to 0$ и $n^2/(2N) \to \infty$,
- центральная область: $0 < c_1 \leqslant n/N \leqslant c_2 < \infty$, где c_1, c_2 — константы,
- правая промежуточная область: $n/N \to \infty$ и $\mathsf{E}\mu_0 \to \infty$,
- правая область: $n/N \to \infty$ и $\mathsf{E}\mu_0 \to \lambda < \infty$, где λ — константа.

Мизес [40; 41] доказал, что в правой области распределение числа пустых ячеек слабо сходится к распределению Пуассона с параметром λ. Бекеши [42] показал, что в левой области распределение случайной величины $\mu_0 - N + n$ слабо сходится к распределению Пуассона с параметром λ. Вайс [43] показал, что в центральной области случайная величина μ_0 асимптотически нормальна. Асимптотическая нормальность случайной величины μ_0 в левой промежуточной и правой промежуточной областях доказана Реньи [44]. Эти результаты дают полное описание всех невырожденных предельных распределений для μ_0 (см. [45]). В работе [46] Ватутиным и Михайловым доказаны аналогичные предельные теоремы для числа пустых ячеек в равновероятной схеме размещения частиц комплектами.

В работе [47] Колчиным доказаны локальные предельные теоремы с оценкой скорости сходимости для μ_0 при значениях, близких к $\mathsf{E}\mu_0$. Эти результаты получены для таких областей, которые перекрываются и охватывают всю область поведения $n, N \to \infty$.

Пусть равновероятно по N ячейкам независимо друг от друга последовательно размещаются частицы. Через $\nu_1(N, k)$ обозначим число испытаний, после которых впервые k ячеек будут содержать хотя бы одну частицу. Так как $\mathsf{P}(\nu_1(N, k) \leqslant n) = \mathsf{P}(\mu_0(n, N) \leqslant N - k)$, то случайные величины $\mu_0(n, N)$ и $\nu_1(N, k)$ тесно связаны между собой. Полное исследование всех предельных распределений $\nu_1(N, k)$ провели Баум и Биллингсли [48].

Предыдущие результаты не позволяют получить асимптотику вероятности $\mathsf{P}(\mu_0 = 0)$, например, в центральной области. В статье [49] при $n = [\theta \cdot N]$, где $\theta \in (0, \infty)$ — константа, доказана предельная теорема интегрального типа для μ_0 в области больших уклонений. В диссертации доказана предельная теорема локального типа для μ_0 в более широкой области. В отличие от локальных предельных теорем, полученных Колчиным [47], расстояние от значения случайной величины μ_0 до $\mathsf{E}\mu_0$ может быть большим.

Цели работы:
1) Исследование распределений межрекордных наполнений.
2) Нахождение предельных распределений максимального расстояния до ближайшего соседа.
3) Исследование распределения числа пустых ячеек в классической схеме случайных размещений в области больших и сверхбольших уклонений.

Научная новизна: Результаты диссертации являются новыми и состоят в следующем:
1) Для новой характеристики в теории рекордов — межрекордных наполнений — доказаны одномерные предельные теоремы.
2) В случае произвольного метрического пространства найдены предельные распределения максимального расстояния до ближайшего соседа.
3) В случае двоичного куба найдено предельное распределение максимального расстояния до ближайшего соседа в широком диапазоне изменения параметров.
4) Доказана локальная предельная теорема для числа пустых ячеек в классической схеме случайных размещений.

Практическая значимость Диссертация имеет теоретический характер. Ее результаты могут быть полезны специалистам в области математической статистики.

Методология и методы исследования. В работе используются аналитические методы теории вероятностей. Предельные теоремы для максимального расстояния до ближайшего соседа в разных случаях получены методом моментов и методом Чена-Стейна.

Апробация работы. Результаты диссертации докладывались автором на семинаре «Дискретные задачи теории вероятностей» кафедры математической статистики и случайных процессов механико-математического факультета МГУ им. М. В. Ломоносова и на следующих конференциях:

1. Международная конференция «Аналитические и вычислительные методы в теории вероятностей и её приложениях», (23–27 октября 2017 г., г. Москва)
2. XV международная конференция «Алгебра, теория чисел и дискретная геометрия: современные проблемы и приложения», (28-31 мая 2018 г., г. Тула)
3. Научная конференция «Ломоносовские чтения – 2016», (апрель 2016 г.)

4. Научная конференция «Ломоносовские чтения – 2018», (апрель 2018 г.)

Публикации. Результаты диссертации опубликованы в четырёх статьях [1—4] в журнале "Дискретная математика" , рекомендованном ВАК.

Краткое содержание диссертации

Во **введении** обоснована актуальность темы диссертации, изложены цели исследования, перечислены основные полученные результаты.

Первая глава.

В *первом параграфе первой главы* вводятся необходимые определения.

Пусть $X_1, X_2, \ldots$ — независимые одинаково распределенные положительные случайные величины, имеющие непрерывную функцию распределения $F(x)$. Последовательность *рекордных моментов* $\{L(k), k \geqslant 1\}$, то есть таких моментов времени t, что случайная величина X_t больше, чем все предыдущие, определяется формулами

$$L(1) = 1, \quad L(k+1) = \min\{t > L(k) : X_t > X_{L_k}\}, \quad k \geqslant 1.$$

При этом значения $\{X(k) = X_{L_k}, k \geqslant 1\}$ называются *рекордными значениями* (или просто *рекордами*). Величины $\Delta(n) = L(n+1) - L(n)$, $n \geqslant 1$, называются *межрекордными интервалами*. Обзоры результатов для рекордов, рекордных моментов и межрекордных интервалов можно найти в [14; 16—18]. Рекордные величины тесно связаны с порядковыми статистиками, в первую очередь, с крайними членами вариационного ряда. Теория экстремальных порядковых статистик достаточно широко изложена, например, в [10—12].

Мы вводим совокупность случайных величин, тесно связанных с рекордами и рекордными моментами. Рассмотрим случайные величины $X_{L(i)+1}, X_{L(i)+2}, \ldots, X_{L(i+1)-1}$ для произвольного фиксированного натурального числа i. Любая из них попадает в один из непересекающихся полуинтервалов $(0; X_{L(1)}]$, $(X_{L(1)}; X_{L(2)}], \ldots,$ $(X_{L(i-1)}; X_{L(i)}]$. Обозначим количество случайных величин, попавших в первый интервал, через α_i^1, количество случайных величин, попавших во второй интервал, через α_i^2 и так далее. Совокупность случайных величин α_i^j, $1 \leqslant j \leqslant i$, назовем *межрекордными наполнениями*:

$$\alpha_i^j = \sum_{l=L(i)+1}^{L(i+1)-1} I\{X_{L(j-1)} < X_l \leqslant X_{L(j)}\}, \, 1 \leqslant j \leqslant i, \tag{1}$$

т.е. α_i^j равно количеству случайных величин в последовательности $\{X_n\}$ с индексами между $L(i)$ и $L(i+1)$ и попавших в промежуток между рекордами $X_{L(j-1)}$ и $X_{L(j)}$. Межрекордные наполнения естественно связаны с рекордными моментами и межрекордными интервалами:

$$\Delta(i) = L(i+1) - L(i) = \sum_{j=1}^{i} \alpha_i^j + 1,$$

$$L(n) = 1 + \sum_{i=1}^{n-1} \Delta(i) = n + \sum_{i=1}^{n-1} \sum_{j=1}^{i} \alpha_i^j,$$

поэтому межрекордные наполнения дают более детальное представление о рекордных моментах и межрекордных интервалах. Изучению распределения межрекордных наполнений посвящена первая глава.

Во *втором параграфе первой главы* представлена формула совместного распределения межрекордных наполнений, сформулированы теоремы о свойствах этого распределения. Результаты этого параграфа получены Н. Ю. Пасынковым в [1] и представлены (без доказательств) только потому, что они относятся к межрекордным наполнениям.

Теорема 1. *Пусть фиксировано $k \in \mathbb{N}$. Для набора целых неотрицательных чисел $\{a_i^j, 1 \leqslant j \leqslant i \leqslant k\}$ обозначим $A_i = \sum\limits_{j=1}^{i} a_i^j$, $A^j = \sum\limits_{i=j}^{k} a_i^j$. Совместное распределение случайных величин α_i^j находится по формуле*

$$\mathsf{P}\left\{\bigcap [\alpha_i^j = a_i^j],\ 1 \leqslant j \leqslant i \leqslant k\right\} = \frac{\prod\limits_{i=1}^{k} A_i! \cdot \prod\limits_{j=1}^{k} A^j!}{N! \cdot \prod\limits_{1 \leqslant j \leqslant i \leqslant k} a_i^j!},$$

где $N = A_1 + \ldots + A_k + k + 1 = A^1 + \ldots + A^k + k + 1$.

Следствие 1 и теоремы 2-4 описывают свойства совместного распределения межрекордных наполнений.

Следствие 1. *Пусть фиксировано $k \in \mathbb{N}$. Тогда для любого $r \in \mathbb{N}$ и любых таких натуральных $i_1, \ldots, i_r$, $j_1, \ldots, j_r$, что $i_1 \leqslant i_2 \leqslant \ldots \leqslant i_r \leqslant k$ и $j_1 \leqslant i_1, \ldots, j_r \leqslant i_r$, распределения наборов*

$$\left(\alpha_{i_1}^{j_1}, \ldots, \alpha_{i_r}^{j_r}\right) \ \text{и} \ \left(\alpha_{k-j_1+1}^{k-i_1+1}, \ldots, \alpha_{k-j_r+1}^{k-i_r+1}\right)$$

одинаковы (симметричность относительно диагонали $\alpha_k^1, \alpha_{k-1}^2, \alpha_{k-2}^3, \ldots$). В частности, α_i^j и α_{i-j+1}^1 одинаково распределены (поэтому одномерные распределения межрекордных наполнений исчерпываются распределениями случайных величин α_k^1, $k \in \mathbb{N}$).

Пусть $m,n \in \mathbb{N}$. Обозначим событие, состоящее в том, что случайные величины α_i^j принимают заданные значения $a_i^j, m \leqslant j \leqslant i \leqslant n$, через $B_m^n = \left\{\cap[\alpha_i^j = a_i^j], m \leqslant j \leqslant i \leqslant n\right\}$.

Теорема 2. *Для любых $m,n \in \mathbb{N}$ и любых чисел $a_i^j \in \mathbb{N}_0$ справедливо равенство*

$$\mathsf{P}\left\{\cap[\alpha_i^j = a_i^j], 1 \leqslant j \leqslant i \leqslant n\right\} = \mathsf{P}\left\{\cap[\alpha_{i+m}^{j+m} = a_i^j], 1 \leqslant j \leqslant i \leqslant n\right\}.$$

Теорема 3. *Для любых $m,n \in \mathbb{N}$ и любых чисел $a_i^j \in \mathbb{N}_0$ справедливо равенство*

$$\mathsf{P}(B_1^m \cap B_{m+1}^{m+n}) = \mathsf{P}(B_1^m) \cdot \mathsf{P}(B_{m+1}^{m+n}).$$

Теорема 4. *Для любых $m,n,q \in \mathbb{N}$ и любых чисел $a_i^j \in \mathbb{N}_0$ справедливо равенство*

$$\mathsf{P}(B_1^{m+q} \cap B_{m+1}^{m+q+n} | B_{m+1}^{m+q}) = \mathsf{P}(B_1^{m+q} | B_{m+1}^{m+q}) \cdot \mathsf{P}(B_{m+1}^{m+q+n} | B_{m+1}^{m+q}).$$

В *третьем параграфе первой главы* представлена интегральная формула для одномерного распределения межрекордного распределения, сформулированы и доказаны предельные теоремы для этого распределения. Приведем полученные результаты для межрекордных наполнений в сравнении с результатами для рекордных моментов и межрекордных интервалов.

В 1962 г. Реньи ([19]) свёл изучение распределения рекордных моментов к изучению распределений сумм независимых случайных величин. Определим случайные индикаторы ξ_k, $k \geqslant 2$, следующим образом: $\xi_k = 1$, если существует такое $n \geqslant 2$, что $L(n) = k$, и $\xi_k = 0$ иначе. Тогда распределение рекордных моментов сводится к распределению сумм независимых случайных величин при помощи равенства

$$\mathsf{P}(L(n) > t) = \mathsf{P}(\xi_1 + \ldots + \xi_t < n).$$

Используя это замечание, Реньи ([19]) получил предельные соотношения

$$\mathsf{P}\left(\frac{\ln L(n) - n}{\sqrt{n}} \leqslant x\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt, \quad \mathsf{P}\left(\lim_{n\to\infty} \frac{\ln L(n)}{n} = 1\right) = 1,$$
$$\mathsf{P}\left(\limsup_{n\to\infty} \frac{\ln L(n) - n}{\sqrt{2n \ln\ln n}} = 1\right) = 1, \quad \mathsf{P}\left(\liminf_{n\to\infty} \frac{\ln L(n) - n}{\sqrt{2n \ln\ln n}} = -1\right) = 1. \quad (2)$$

Асимптотическое поведение межрекордных интервалов $\Delta(n)$ впервые исследовал Ньютс ([50]) с помощью представления распределения $\Delta(n)$:

$$\mathsf{P}(\Delta(n) > m) = \int_0^\infty \frac{x^{n-1}}{(n-1)!} e^{-x} (1 - e^x)^m dx, \quad m = 1, 2, \ldots \quad (3)$$

Используя равенство (3), Ньютс ([50]) получил предельные соотношения

$$\sqrt[n]{\Delta(n)} \xrightarrow[n\to\infty]{\mathsf{P}} e, \quad \mathsf{P}\left(\frac{\ln\Delta(n)-n}{\sqrt{n}} \leqslant x\right) \xrightarrow[n\to\infty]{} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}dt. \tag{4}$$

Позже, с помощью того же равенства (3), в [51] получено предельное соотношение

$$\mathsf{P}\left(\frac{\ln\Delta(n)}{n} \to 1\right) = 1, \tag{5}$$

и в [52] — равенства

$$\mathsf{P}\left(\limsup_{n\to\infty}\frac{\ln\Delta(n)-n}{\sqrt{2n\ln\ln n}} = 1\right) = \mathsf{P}\left(\liminf_{n\to\infty}\frac{\ln\Delta(n)-n}{\sqrt{2n\ln\ln n}} = -1\right) = 1. \tag{6}$$

Позже в статье [53] было доказано, что

$$\mathsf{P}\left(\limsup_{n\to\infty}\left|\frac{\ln\Delta(n)-\ln L(n)}{\ln n}\right| = 1\right) = 1. \tag{7}$$

Из (7) и (2) следуют все предельные соотношения (4), (5), (6).

Автор получил представление распределения α_i^j, $i > j \geqslant 1$, аналогичное представлению (3).

Теорема 5. *Для любого натурального m и таких натуральных i, j, что $i > j$, распределение случайной величины α_i^j записывается в виде*

$$\mathsf{P}(\alpha_i^j \geqslant m) = \int_0^\infty \frac{x^{i-j-1}}{(i-j-1)!}e^{-x}\int_0^\infty\left[1-\frac{e^{-x}}{e^y+e^{-x}-1}\right]^m e^{-y}\,dy\,dx.$$

Автором доказано также утверждение, аналогичное утверждению (7):

$$\mathsf{P}\left(\limsup_{n\to\infty}\left|\frac{\ln\alpha_n^1-\ln\Delta(n)}{\ln n}\right| \leqslant 1\right) = 1. \tag{8}$$

Из (8), (4), (5) и (6) получаем следующую теорему.

Теорема 6. *Справедливы предельные соотношения:*

$$\frac{\ln\alpha_i^j-(i-j+1)}{\sqrt{i-j+1}} \xrightarrow[i-j\to\infty]{D} N(0,1), \qquad \lim_{i-j\to\infty}\frac{\ln\alpha_i^j}{i-j+1} \overset{п.н.}{=} 1,$$

$$\limsup_{i-j\to\infty}\frac{\ln\alpha_i^j-(i-j+1)}{2(i-j+1)\ln\ln(i-j+1)} \overset{п.н.}{=} 1,$$

$$\liminf_{i-j\to\infty}\frac{\ln\alpha_i^j-(i-j+1)}{2(i-j+1)\ln\ln(i-j+1)} \overset{п.н.}{=} -1.$$

В работе Таты [20] показано, что

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{L(n)}{L(n-1)} > x\right) = \frac{1}{x}, \quad x \geqslant 1. \tag{9}$$

Соотношение (9) эквивалентно соотношению

$$\lim_{n\to\infty} \mathsf{P}\left(\frac{\Delta(n-1)}{L(n)} \leqslant x\right) = x, \quad x \in [0,1].$$

Автором доказана следующая теорема.

Теорема 7. *Справедливо предельное соотношение:*

$$\mathsf{P}\left\{\frac{\alpha_k^1}{\Delta(k)} \leqslant x\right\} \xrightarrow[k\to\infty]{} x, \qquad 0 \leqslant x \leqslant 1.$$

Вторая глава.

Вторая глава диссертации посвящена изучению предельных распределений экстремальных расстояний до ближайшего соседа.

Случайные величины, связанные с расстояниями до ближайших соседей (как правило, для многомерных торов), изучались многими авторами (см., например, [36—39]). В книге [54] рассмотрен случай многомерного компактного однородного риманова многообразия (примерами являются многомерные сфера и тор).

В диссертации доказана предельная теорема для максимального расстояния до ближайшего соседа в произвольном метрическом пространстве (B, ρ). Главное отличие от перечисленных работ заключается в том, что пространство B может быть дискретным (например, двоичным кубом). Случай двоичного куба рассмотрен более детально. Во многих работах изучались задачи, связанные с близкими подстроками в последовательностях символов конечного алфавита (см., например, [55—57]), а также с предельным случаем точных повторений подстрок (см., например, [58—62]).

Расстояния до ближайших соседей используются, например, при построении статистических критериев равномерности распределения (см., например, [29—32]), при оценке энтропии (см., например, [33; 34]), в алгоритмах классификации (см., например [35]).

В *первом параграфе второй главы* вводятся обозначения.

Пусть на метрическом пространстве (B, ρ) задана вероятностная мера Q, удовлетворяющая условию

$$Q\left(\{y \in B : \rho(y,x) \leqslant z\}\right) = w(z) \quad \text{для любого} \quad x \in B, \tag{10}$$

и $\xi_1, \ldots, \xi_n$ — независимые случайные элементы со значениями в B с распределением Q, удовлетворяющим условию (10). Естественными примерами таких мер являются равномерные распределения на многомерных сферах и (непрерывных или дискретных) торах.

Для каждого $i \in \{1, \ldots, n\}$ введём случайную величину $\zeta_i = \min_{j \neq i} \rho(\xi_i, \xi_j)$ — расстояние от точки ξ_i до ее ближайшего соседа. Пусть $\zeta_{(1)} \leqslant \ldots \leqslant \zeta_{(n)}$ — вариационный ряд, составленный из величин $\zeta_1, \ldots, \zeta_n$. Тогда величина $\phi_n = \zeta_{(1)} = \min_{1 \leqslant i < j \leqslant n} \rho(\xi_i, \xi_j)$ является минимальным расстоянием между точками выборки (минимальным расстоянием до ближайшего соседа), а величина $\psi_n = \zeta_{(n)} = \max_{1 \leqslant i \leqslant n} \zeta_i$ является максимальным расстоянием до ближайшего соседа.

Во *втором параграфе* второй главы методом Чена-Стейна доказана предельная теорема для распределения минимального расстояния между точками выборки.

Теорема 10. *Если r_n — такая последовательность чисел, что $C_n^2 w(r_n) \to \lambda \in (0, +\infty)$ при $n \to \infty$, то*

$$\mathsf{P}(\phi_n > r_n) \to e^{-\lambda}, \quad n \to \infty.$$

В статье [2] эта теорема была доказана А. М. Зубковым методом моментов.

Третий параграф второй главы посвящен максимальному расстоянию до ближайшего соседа. Доказаны теоремы о предельных распределениях максимального расстояния до ближайшего соседа. В качестве примеров рассмотрены многомерный тор и двоичный куб.

В *первом подпараграфе* методом моментов доказана теорема о предельном распределении максимального расстояния до ближайшего соседа в произвольном метрическом пространстве.

В формулировке теоремы используются величины

$$L_k(z) = \min_{\substack{x_1, \ldots, x_k \in B \\ \rho(x_i, x_j) > z \\ 1 \leqslant i < j \leqslant k}} Q\left(\bigcup_{i=1}^{k} \{y \in B : \rho(y, x_i) \leqslant z\} \right), \quad k \geqslant 1, z \geqslant 0, \tag{11}$$

равные минимальной вероятностной мере объединения k таких шаров радиуса z, что центр каждого шара не принадлежит другим шарам.

Теорема 11. *Если выполнены условия*
1) $n(1 - w(r_n))^n \xrightarrow[n \to \infty]{} \lambda, \quad \lambda \in (0, \infty)$,
2) $w(2r_n) \xrightarrow[n \to \infty]{} 0$,
3) *при любых фиксированных натуральных* $1 \leqslant k < m$

$$n^m (w(2r_n) - w(r_n))^{m-k} \max_{i_1 + \ldots + i_k = m} (1 - L_{i_1}(r_n) - \ldots - L_{i_k}(r_n))^n \xrightarrow[n \to \infty]{} 0,$$

то

$$\mathsf{P}(\psi_n \leqslant r_n) \xrightarrow[n\to\infty]{} e^{-\lambda}.$$

Во *втором подпараграфе* рассмотрен случай многомерного тора размерности d. Проверены условия теорем 10, 11 и, соответственно, доказана следующая теорема.

Теорема 12. *При $d = \mathrm{const}, n \to \infty$ справедливы соотношения:*

$$\mathsf{P}\left(\phi_n > \frac{1}{2}\left(\frac{2y}{n^2}\right)^{1/d}\right) \to e^{-y},\ y > 0;\ \mathsf{P}\left(\psi_n \leqslant \left(\frac{y+\ln n}{2^d n}\right)^{1/d}\right) \to e^{-e^{-y}},\ y \in \mathbb{R}.$$

В *третьем подпараграфе* рассмотрен случай двоичного куба $\{0,1\}^T$ при определенном изменении параметров. Доказана следующая теорема.

Теорема 13. *Если n,$T \to \infty$ и s_n,r_n меняются так, что*

$$C_n^2 \frac{1}{2^T} \sum\nolimits_{k=0}^{s_n} C_T^k \to \lambda,$$

$$n\left(1 - \frac{1}{2^T}\sum\nolimits_{k=0}^{r_n} C_T^k\right)^n \to \lambda \in (0,\infty), \quad r_n = o(T),$$

то

$$\mathsf{P}(\phi_n > s_n) \to e^{-\lambda}, \quad \mathsf{P}(\psi_n \leqslant r_n) \to e^{-\lambda}.$$

В *четвертом подпараграфе* методом Чена-Стейна доказана теорема о предельном распределении максимального расстояния до ближайшего соседа. Улучшены условия теоремы 11.

Теорема 15. *Пусть при $n \to \infty$ пространства $B = B_n$ и меры $Q = Q_n$ удовлетворяют условию (10), а последовательность r_n такова, что выполнены условия*
1) $n(1 - w(r_n))^n \xrightarrow[n\to\infty]{} \lambda \in (0, +\infty)$,
2) $n^2(w(2r_n) - w(r_n))(1 - L_2(r_n))^n \xrightarrow[n\to\infty]{} 0$.
Тогда

$$\mathsf{P}(\psi_n \leqslant r_n) \xrightarrow[n\to\infty]{} e^{-\lambda}.$$

В *пятом подпараграфе* методом моментов доказана теорема о предельном распределении максимального расстояния до ближайшего соседа в случае двоичного куба при изменении параметров, не рассматривавшемся в теореме 13.

Теорема 16. *Пусть $n, T \to \infty$. Если последовательность целых чисел r_n такова, что выполнены условия:*
1) $n\left(1 - \frac{1}{2^T}\sum_{k=0}^{r_n} C_T^k\right)^n \to \lambda \in (0, +\infty)$,

2) $r_n/T \to \alpha \in [0, 1/2)$,
то

$$\mathsf{P}(\psi_n \leqslant r_n) \to e^{-\lambda}, \tag{12}$$

при этом

$$\mathsf{P}(\psi_n \leqslant r_n - 1) \to 0, \quad \mathsf{P}(\psi_n \leqslant r_n + 1) \to 1.$$

Третья глава.

В *первом параграфе третьей главы* сформулирован полученный результат. Пусть $\mu_0(n, N)$ — число пустых ячеек при случайном равновероятном размещении n частиц по N ячейкам. Легко вычислить, что $\mathsf{E}\mu_0 = N(1 - 1/N)^n$. При $n, N \to \infty$ выделяют пять областей, связывающих n и N:

- левая область: $n^2/(2N) \to \lambda < \infty$, где λ — константа,
- левая промежуточная область: $n/N \to 0$ и $n^2/(2N) \to \infty$,
- центральная область: $0 < c_1 \leqslant n/N \leqslant c_2 < \infty$, где c_1, c_2 — константы,
- правая промежуточная область: $n/N \to \infty$ и $\mathsf{E}\mu_0 \to \infty$,
- правая область: $n/N \to \infty$ и $\mathsf{E}\mu_0 \to \lambda < \infty$, где λ — константа.

В центральной, левой промежуточной и правой промежуточной областях число пустых ячеек асимптотически нормально. В правой области распредление числа пустых ячеек слабо сходится к распределению Пуассона с параметром λ. В левой области распределение случайной величины $\mu_0 - N + n$ слабо сходится к распределению Пуассона с параметром λ. Эти результаты дают полное описание всех невырожденных предельных распределений для μ_0 (см. [45]).

Из этих результатов нельзя получить асимптотику вероятности $\mathsf{P}(\mu_0 = 0)$, например, в центральной области. В статье [49] при $n = [\theta \cdot N]$, где $\theta \in (0, \infty)$ — константа, доказана предельная теорема интегрального типа для μ_0 в области больших уклонений.

Результатом этой главы является предельная теорема для μ_0 локального типа в более широкой области.

Теорема 18. *Пусть $k \geqslant 0$ и $n, N \to \infty$, причем $1 < c_1 \leqslant n/(N-k) \leqslant c_2 < \infty$, где c_1, c_2 — константы. Тогда верна асимптотика*

$$\mathsf{P}(\mu_0 = k) = C_N^k \left(1 - \frac{k}{N}\right)^n e^{n(y_0 - 1)} y_0^{N-n-k} \sqrt{\frac{N-k}{N-k+n(y_0-1)}}(1 + o(1)), \tag{13}$$

где y_0 — положительный корень уравнения

$$1 - e^{-yn/(N-k)} = y. \tag{14}$$

Замечание. Если $n/(N-k) < 1$, то, очевидно, $\mathsf{P}(\mu_0 = k) = 0$.

В правой и правой промежуточной областях $n/N \to \infty$, поэтому теорема 18 не применима. Теорему 18 можно использовать в центральной, левой, левой промежуточной областях.

Следствие 2. *В левой и левой промежуточной областях для любого фиксированного* $\gamma \in (0,1)$ *верна асимптотика*

$$\mathsf{P}(\mu_0 = N - [\gamma n]) = C_N^{[\gamma n]} \left(\frac{[\gamma n]}{N}\right)^n e^{\frac{[\gamma n](y_0-1)}{n}} y_0^{[\gamma n]-n} \sqrt{\frac{\gamma}{\gamma + y_0 - 1}}(1+o(1)),$$

где y_0 *— положительный корень уравнения*

$$1 - e^{-y/\gamma} = y.$$

Следствие 3. *Если в центральной области* $0 < c_1 \leqslant n/N \leqslant c_2 < \infty$ *и* $(1 - c_1/\gamma_1)N \leqslant k \leqslant N/\gamma_2$, *где* $c_1, c_2, \gamma_1, \gamma_2$ *— константы, причем* $\gamma_1, \gamma_2 > 1$, *то верна асимптотика* (13), *где* y_0 *— положительный корень уравнения* (14).

Следствие 4. *Если* $n, N \to \infty$, *причем* $1 < c_1 \leqslant n/N \leqslant c_2 < \infty$, *где* c_1, c_2 *— константы, то верна асимптотика*

$$\mathsf{P}(\mu_0 = 0) = \sqrt{\frac{N}{N + n(y_0 - 1)}} e^{n(y_0-1)} y_0^{N-n} \cdot (1 + o(1)), \tag{15}$$

где y_0 *— положительный корень уравнения*

$$1 - e^{-ny/N} = y.$$

Во *втором параграфе третьей главы* доказана теорема 18.

Благодарности. Автор благодарит своих научных руководителей В. В. Козлова за постановку задачи главы 1 и А. М. Зубкова за постановку задач глав 2 и 3, за ценные советы и постоянное внимание к работе.

Публикации автора по теме диссертации

1. *Орлов О. П.*, *Пасынков Н. Ю.* Распределения межрекордных наполнений // Дискретная математика. — 2015. — Т. 27, № 3. — С. 56—73.
2. *Зубков А. М.*, *Орлов О. П.* Предельные распределения экстремальных расстояний до ближайшего соседа // Дискретная математика. — 2017. — Т. 29, № 2. — С. 3—17.
3. *Орлов О. П.* Предельные распределения максимального расстояния до ближайшего соседа // Дискретная математика. — 2018. — Т. 30, № 3. — С. 88—98.

4. *Орлов О. П.* Локальная предельная теорема для числа пустых ячеек при случайных равновероятных размещениях // Дискретная математика. — 2021. — Т. 33, № 4. — С. 83—93.

Список литературы

5. *Chandler K. N.* The distribution and frequency of record values // Journal of the Royal Statistical Society: Series B (Methodological). — 1952. — Vol. 14, no. 2. — P. 220—228.

6. *Foster F., Stuart A.* Distribution-Free Tests in Time-Series Based on the Breaking of Records Division of Research Techniques, London School of Economics // Journal of the Royal Statistical Society: Series B (Methodological). — 1954. — Vol. 16, no. 1. — P. 1—13.

7. *Foster F., Teichroew D.* A sampling experiment on the powers of the records tests for trend in a time series // Journal of the Royal Statistical Society. Series B (Methodological). — 1955. — P. 115—121.

8. *Stuart A.* The efficiencies of tests of randomness against normal regression // Journal of the American Statistical Association. — 1956. — Vol. 51, no. 274. — P. 285—287.

9. *Stuart A.* The efficiency of the records test for trend in normal regression // Journal of the Royal Statistical Society: Series B (Methodological). — 1957. — Vol. 19, no. 1. — P. 149—153.

10. *Галамбош Я.* Асимптотическая теория экстремальных порядковых статистик. — Москва : Наука, 1984. — 304 с.

11. *Дэйвид Г.* Порядковые статистики. — М. : Наука, 1979. — 336 с.

12. *Ahsanullah M., Nevzorov V. B., Shakil M.* An introduction to order statistics. — Amsterdam - Paris - Beijing : ATLANTIS PRESS, 2013. — 244 p.

13. *Arnold B. C., Balakrishnan N., Nagaraja H. N.* Records. — New York, Chichester, Weinheim, Brisbane, Singapore, Toronto : JOHN WILEY & SONS, 1998. — 312 p.

14. *Невзоров В. Б.* Рекорды. Математическая теория. — Москва : Фазис, 2000. — 256 с.

15. *Ahsanullah M., Nevzorov V. B.* Records via Probability Theory. — Amsterdam - Paris - Beijing : ATLANTIS PRESS, 2015. — 255 p.

16. *Glick N.* Breaking records and breaking boards // The American Mathematical Monthly. — 1978. — Vol. 85, no. 1. — P. 2—26.

17. *Невзоров В. Б.* Рекорды // Теория вероятн. и ее примен. — 1987. — Т. 32, № 2. — С. 219—251.

18. *Stepanov A.* On the Mathematical Theory of Records // Communications in Mathematics. — 2021. — Vol. 29, no. 1. — P. 151—162.

19. *Renyi A.* On the extreme elements of observations // Selected papers of Alfred Renyi. — Budapest: Akademiai Kiado, 1976. — Vol. 3. — P. 50—65.

20. *Tata M. N.* On outstanding values in a sequence of random variables // Z.Wahrscheinlichkeitheor. verw. Geb. — 1969. — Vol. 12, no. 1. — P. 9—20.

21. *Shorrock R. W.* On record values and record times // J. Appl. Probab. — 1972. — Vol. 9, no. 2. — P. 316—326.

22. *Shorrock R. W.* A limit theorem for inter-record times // J. Appl. Probab. — 1972. — Vol. 9, no. 1. — P. 219—223.

23. *Shorrock R. W.* Record values and inter-record times // J. Appl. Probab. — 1973. — Vol. 10, no. 3. — P. 543—555.

24. *Adam M. B., Tawn J. A.* Bivariate extreme analysis of Olympic swimming data // Journal of Statistical Theory and Practice. — 2012. — Vol. 6, no. 3. — P. 510—523.

25. *Adam M., Tawn J. A.* Modelling record times in sport with extreme value methods // Malaysian Journal of Mathematical Sciences. — 2016. — Vol. 10, no. 1. — P. 1—21.

26. *Hashorva E.* On the number of near-maximum insurance claim under dependence // Insurance: Mathematics and Economics. — 2003. — Vol. 32, no. 1. — P. 37—49.

27. *Hashorva E., Hüsler J.* Estimation of tails and related quantities using the number of near-extremes // Communications in Statistics-Theory and Methods. — 2005. — Vol. 34, no. 2. — P. 337—349.

28. *Ramesh S. V. T., Salas J. D., Stedinger J. R.* Statistical Analysis of Hydrologic Variables: Methods and Applications. — American Society of Civil Engineers, 2019. — 548 p.

29. *Bickel P. J., Breiman L.* Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test // Ann. Probab. — 1983. — Vol. 11, no. 1. — P. 185—214.

30. *Shilling M. F.* Goodness of fit testing in R m based on the weighted empirical distribution of certain nearest neighbor statistics // Ann. Statist. — 1983. — Vol. 11, no. 1. — P. 1—12.

31. *Shilling M. F.* An infinite-dimensional approximation for nearest neighbor goodness of fit tests // Ann. Statist. — 1983. — Vol. 11, no. 1. — P. 13—24.

32. *L'écuyer P.*, *Cordeau J. F.*, *Simard R.* Close-point spatial tests and their application to random number generators // Oper. Res. — 2000. — Vol. 48, no. 2. — P. 308—317.

33. Nonparametric entropy estimation: An overview / J. Beirlant [et al.] // International Journal of Mathematical and Statistical Sciences. — 1997. — Vol. 6, no. 1. — P. 17—39.

34. *Leonenko N.*, *Pronzato L.*, *Savani V.* A class of Rényi information estimators for multidimensional densities // The Annals of Statistics. — 2008. — Vol. 36, no. 5. — P. 2153—2182.

35. *Kumar R. R.*, *Viswanath P.*, *Bindu C. S.* Nearest neighbor classifiers: a review // Int. J. Comput. Intell. Res. — 2017. — Vol. 13, no. 2. — P. 303—311.

36. *Silverman B.*, *Brown T.* Short distances, flat triangles and Poisson limits // J. Appl. Probab. — 1978. — Vol. 15. — P. 815—825.

37. *Henze N.* The limit distribution for maxima of «weighted» r-th nearest-neighbor distances // J. Appl. Prob. — 1982. — Vol. 19, no. 2. — P. 344—354.

38. *Penrose M. D.*, *Yukich J. E.* Laws of large numbers and nearest neighbor distances // Advances in Directional and Linear Statistics. — Berlin: Physica-Verlag HD, 2011. — P. 189—199.

39. *Baryshnikov Y.*, *Penrose M. D.*, *Yukich J. E.* Gaussian limits for generalized spacings // Ann. Appl. Probab. — 2009. — Vol. 19, no. 1. — P. 158—185.

40. *Mises R.* Ueber Aufteilungs und Besetzungs-wahrscheinlichkeiten // Revue de la Faculte des Sciences de l'Universite d'Istanbul. — 1939. — T. 4. — C. 145—163.

41. Selected Papers of Richard von Mises / P. Frank [и др.]. — Providence : American Mathematical Society, 1964. — 568 с.

42. *Bekessy A.* On classical occupancy problem, I. // Magy. tud. akad. Mat. kutato int. kozl. — 1963. — Vol. 8, no. 1/2. — P. 59—71.

43. *Weiss I.* Limiting distributions in some occupancy problems // The Annals of Mathematical Statistics. — 1958. — Vol. 29, no. 3. — P. 878—884.

44. *Rényi A.* Three new proofs and a generalization of a theorem of Irving Weiss // Magyar Tud. Akad. Mat. Kutató Int. Közl. — 1962. — Vol. 7, no. 1/2. — P. 203—214.

45. *Колчин В. Ф.*, *Севастьянов Б. А.*, *Чистяков В. П.* Случайные размещения. — Москва : Наука, 1976. — 223 с.

46. *Ватутин В. А.*, *Михайлов В. Г.* Предельные теоремы для числа пустых ячеек в равновероятной схеме размещения частиц комплектами // Теория вероятностей и ее применения. — 1982. — Т. 27, № 4. — С. 684—692.

47. *Колчин В. Ф.* Скорость приближения к предельным распределениям в классической задаче о дробинках // Теория вероятн. и ее примен. — 1966. — Т. 11, № 1. — С. 144—156.

48. *Baum L. E.*, *Billingsley P.* Asymptotic distributions for the coupon collector's problem // The Annals of Mathematical Statistics. — 1965. — Vol. 36, no. 6. — P. 1835—1839.

49. *Dupuis P.*, *Zhang J.*, *Whiting P.* Refined Large Deviation Asymptotics for the Classical Occupancy Problem // Methodol. Comput. Appl. Probab. — 2006. — Vol. 8. — P. 467—496.

50. *Neuts M. F.* Waiting times between record observations // J. Appl. Probab. — 1967. — Vol. 4, no. 1. — P. 206—208.

51. *Holmes P. T.*, *Strawderman W.* A note on the waiting times between record observations // J. Appl. Probab. — 1969. — Vol. 6, no. 3. — P. 711—714.

52. *Strawderman W.*, *Holmes P. T.* On the law of the iterated logarithm for inter-record times // J. Appl. Probab. — 1970. — Vol. 7, no. 2. — P. 432—439.

53. *Galambosh J.*, *Seneta E.* Record times // Proc. Amer. Math. Soc. — 1975. — Vol. 50. — P. 383—387.

54. *Barbour A. D.*, *Holst L.*, *Janson S.* Poisson approximation. — Oxford : Clarendon press, 1992. — vi + 277.

55. *Михайлов В. Г.* Центральная предельная теорема для числа неполных длинных повторений // Теория вероятн. и ее примен. — 1975. — Т. 20, № 4. — С. 880—884.

56. *Михайлов В. Г.* Предельные теоремы пуассоновского типа для числа неполных совпадений s-цепочек // Теория вероятн. и ее примен. — 2002. — Т. 47, № 2. — С. 350—357.

57. *Burden C. J.*, *Kantorovitz M. R.*, *Wilson S. R.* Approximate word matches between two random sequences // Ann. Appl. Probab. — 2008. — Vol. 18, no. 1. — P. 1—21.

58. *Зубков А. М.*, *Михайлов В. Г.* Предельные распределения случайных величин, связанных с длинными повторениями в последовательности независимых испытаний // Теория вероят. и ее примен. — 1974. — Т. 19, № 1. — С. 173—181.

59. *Karlin S.*, *Ost F.* Counts of long aligned word matches among random letter sequences // Adv. Appl. Prob. — 1987. — Vol. 19, no. 2. — P. 293—351.

60. *Зубков А. М.*, *Круглов В. И.* Повторения цепочек на бинарном дереве со случайными метками вершин // Дискретная математика. — 2015. — Т. 27, № 4. — С. 38—48.

61. *Михайлов В. Г.* Оценки точности пуассоновской аппроксимации для распределения числа серий повторений длинных цепочек в цепи Маркова // Дискретная математика. — 2015. — Т. 27, № 4. — С. 67—78.

62. *Михайлов В. Г.* О вероятности наличия в случайной последовательности цепочек с одинаковой структурой // Дискретная математика. — 2016. — Т. 28, № 3. — С. 97—110.

Орлов Олег Павлович

Комбинаторно-вероятностные задачи для случайных выборок

Автореф. дис. на соискание ученой степени канд. физ.-мат. наук

Подписано в печать 22.09.2022 г.
Тираж 100 экз.

Отпечатано в типографии «Реглет»
г. Москва, пр-т Вернадского, д. 39
8 (495) 979-98-99, www.reglet.ru