На правах рукописи

Черепанова Елена Владимировна

ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ ДЛЯ НЕКОТОРЫХ СЛУЧАЙНЫХ КОМБИНАТОРНЫХ СТРУКТУР

01.01.09 — дискретная математика
и математическая кибернетика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Петрозаводск 2004

Работа выполнена в Институте прикладных математических исследований Карельского научного центра РАН.

Научный руководитель:
доктор физико-математических наук, профессор Ю. Л. Павлов.

Официальные оппоненты:
доктор физико-математических наук, профессор В. Ф. Колчин,
кандидат физико-математических наук, доцент А. В. Соколов.

Ведущая организация:
Петрозаводский государственный университет.

Защита состоится 22 июня 2004 г. в 16 часов 15 мин. на заседании диссертационного совета К 002.142.01 в Институте прикладных математических исследований Карельского научного центра РАН по адресу: 185610, Петрозаводск, ул. Пушкинская, И.

С диссертацией можно ознакомиться в библиотеке Карельского научного центра РАН.

Автореферат разослан "12" мая 2004 г.

Ученый секретарь
диссертационного совета
К 002.142.01
к.ф.-м.н., доцент

В. Т. Вдовицын.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. В настоящее время в комбинаторном анализе широко применяются хорошо развитые в теории вероятностей методы. Возможность применения вероятностных методов при решении комбинаторных задач обеспечивается заданием распределения вероятностей на множестве изучаемых комбинаторных объектов, поскольку в этом случае их числовые характеристики можно рассматривать как случайные величины. При этом, задание равномерного распределения позволяет исключить из рассмотрения ту небольшую часть исследуемых объектов, которые обладают нетипичными свойствами. В современных работах по дискретной математике значительное внимание уделяется изучению случайных комбинаторных структур. Примерами таких структур являются случайные графы, случайные подстановки, урновые схемы, представляющие собой удобное средство моделирования разнообразных объектов. Изучение предельных свойств комбинаторных структур, проявляющихся при неограниченном росте числа элементов этих структур, является одним из важнейших направлений исследований. Результаты, полученные для случайных деревьев, лесов, графов подстановок, находят применение в анализе вычислительных алгоритмов, статистических методах, моделировании транспортных и информационных систем, криптографии, теории случайных уравнений, исследовании эволюции случайных графов.

Цель исследования. Целью диссертации является получение предельных теорем для важнейших характеристик случайных подстановок с известным числом циклов, лесов Гальтона — Ватсона и урновой схемы с разноцветными шарами. В частности, предполагалось получение полного описания предельного поведения числа циклов заданной длины случайной подстановки с известным числом циклов, минимальной длины цикла такой подстановки и минимального объема дерева в лесе Гальтона — Ватсона, а также изучение предельного поведения числа пар в обобщенной схеме размещения.

Объекты исследования. Объектами изучения в диссертации являются случайные подстановки с известным числом циклов и случайные рекурсивные леса, леса Гальтона — Ватсона, а также урновая схема с разноцветными шарами.

Методы исследования. Методами исследования в диссертации являются обобщенная схема размещения, методы получения локальных предельных теорем для сумм независимых решетчатых случайных величин, методы теории ветвящихся процессов, метод характеристических функций, алгебраические методы, методы комбинаторики, методы проверки статистических гипотез. Хотя в диссертации рассматриваются достаточно разные случайные комбинаторные структуры, проведенные исследования объединяет основной метод получения предельных теорем — обобщенная схема размещения. Этот метод введен и исследован В. Ф. Колчиным и является одним из наиболее известных и успешно применяемых вероятностных методов решения комбинаторных задач. Использование обобщенной схемы размещения приводит к необходимости доказательства предельных теорем для серий сумм независимых целочисленных случайных величин. Доказательства таких теорем, включая двумерный случай, составили основную сложность при получении результатов диссертации.

Научная новизна. В диссертации впервые получено полное описание предельного поведения числа циклов заданной длины и минимальной длины цикла случайной подстановки с известным числом циклов, а также минимального объема дерева в лесе Гальтона — Ватсона. Доказаны предельные теоремы для числа пар извлеченных шаров одинакового цвета в урновой схеме с разноцветными шарами для не рассматривавшейся ранее области изменения параметров, а также для числа цепей в случайном корневом лесе с помеченными вершинами и для числа путей в графе случайной подстановки с известным числом циклов. Все полученные результаты являются новыми.

Основные положения диссертации, выносимые на защиту. На защиту выносятся:

1. Полное описание предельного поведения числа циклов заданной длины в случайной подстановке с известным числом циклов.

2. Предельные распределения минимальной длины цикла случайной подстановки с известным числом циклов и минимального объема дерева в лесе Гальтона — Ватсона.

3. Предельные распределения числа пар шаров в урновой схеме с разноцветными шарами, числа путей в графе случайной подстановки с известным числом циклов, числа цепей в случайном корневом лесе с помеченными вершинами.

4. Асимптотика статистики типа χ^2, предназначенной для проверки гипотез о распределении вероятностей в некоторых задачах, связанных с урновой схемой, случайными подстановками и случайными лесами.

5. Показана применимость критерия пустых ящиков для проверки статистической гипотезы о равномерности распределения вероятностей на множестве лесов Гальтона — Ватсона.

Связь работы с крупными научными программами, темами. Результаты диссертации были получены в рамках темы плана научно-исследовательских работ Института прикладных математических исследований Карельского научного центра РАН "Вероятности на деревьях и лесах", №гос. регистрации 01.2.00 1 03997. В 2001-2002г. работа выполнялась при поддержке Российского Фонда Фундаментальных Исследований (грант 00-01-00233). В 2003г. исследования проводились по государственному контракту "Исследование случайных комбинаторных структур" по программе фундаментальных исследований РАН "Современные проблемы теоретической математики" (№100002-251/ОМН-01/018-026/090703-1029) и при поддержке гранта НШ 1758.2003.1 Президента Российской Федерации для государственной поддержки ведущих научных школ Российской Федерации (руководитель школы — академик Ю. В. Прохоров).

Апробация результатов диссертации. Основные результаты докладывались на Workshop "Networking games and resource allocation"(Петрозаводск, 2002), Kalashnikov Memorial Seminar (Петрозаводск, 2002), III и IV Всероссийских симпозиумах по прикладной и промышленной математике (Сочи, 2002, Петрозаводск, 2003), X Всероссийской школе-коллоквиуме по стохастическим методам (Сочи, 2003).

Публикация результатов. Основные результаты диссертации опубликованы в девяти работах, из них две статьи в журнале "Дискретная математика", две статьи в сборнике трудов Института прикладных математических исследований Карельского научного центра РАН, пять тезисов докладов на международных, всероссийских и региональных конференциях.

Структура и объем диссертации. Диссертация состоит из введения, пяти глав и списка литературы. Общий объем диссертации составляет 137 страниц. Список литературы содержит 70 наименований.

СОДЕРЖАНИЕ РАБОТЫ

Во введении приводятся обоснование актуальности темы диссертации, формулировка цели работы, основные научные положения, выносимые на защиту, и описание структуры диссертации.

Первая глава носит вспомогательный характер. В ней приводятся определение и основные свойства обобщенной схемы размещения, примеры сведения комбинаторных задач к обобщенной схеме размещения. В примерах и в третьем разделе этой главы содержится описание случайных комбинаторных структур, числовые характеристики которых изучаются в диссертации.

Обобщенная схема размещения задается следующим образом. Пусть имеется набор целочисленных неотрицательных случайных величин $\eta_1,\ldots,\eta_N$ таких, что $\eta_1+\ldots+\eta_N=n$. Эти случайные величины можно рассматривать как заполнения ячеек при размещении n частиц в N ячеек, занумерованных числами $1,2,\ldots,N$, то есть η_i — число частиц, попавших в i-ую ячейку, $1\leqslant i\leqslant N$. Если совместное распределение случайных величин $\eta_1,\ldots,\eta_N$ можно представить в виде

$$\begin{aligned}&\mathbf{P}\{\eta_1=k_1,\ldots,\eta_N=k_N\}=\\&=\mathbf{P}\{\xi_1=k_1,\ldots,\xi_N=k_N\,|\,\xi_1+\ldots+\xi_N=n\},\end{aligned}\tag{1}$$

где $k_1,\ldots,k_N$ — произвольные целые числа, $\xi_1,\ldots,\xi_N$ — независимые неотрицательные целочисленные случайные величины, то говорят, что случайные величины $\eta_1,\ldots,\eta_N$ и $\xi_1,\ldots,\xi_N$ образуют обобщенную схему размещения.

Такая схема размещения является обобщением так называемой классической задачи о размещении, в которой распределение различных частиц по ячейкам происходит равновероятно.

Подстановку можно представить в виде ориентированного графа, состоящего из компонент связности, которые являются циклами и называются циклами подстановки. В диссертации рассматривается множество $S_{n,N}$ подстановок степени n, имеющих N циклов, при равномерном распределении вероятностей на этом множестве, и изучается предельное поведение числа циклов заданной длины, минимальной длины цикла, а также числа путей в графах таких случайных подстановок.

Рекурсивный лес — это лес с занумерованными вершинами, который строится так, что очередная, i-ая, вершина либо присоединяется к одной из предыдущих $i-1$ вершин леса, либо становится первой вершиной дерева, еще не имеющего вершин. В работе рассматривается множество $R_{N,n}$ рекурсивных лесов, состоящих из N некорневых деревьев и имеющих n вершин, на котором задано равномерное распределение вероятностей.

Рассмотрим урновую схему с разноцветными шарами. Пусть из урны, содержащей по $m \geqslant 2$ различных шаров каждого N цветов, осуществляется равновероятный выбор без возвращения n шаров. В диссертации изучается предельное поведение числа пар извлеченных шаров одинакового цвета в этой урновой схеме при $n, N \to \infty$.

Обозначим через $\mathfrak{F}_N$ лес Гальтона — Ватсона, образованный ветвящимся процессом с N начальными частицами, в котором число прямых потомков каждой частицы задается случайной величиной ξ с распределением

$$\mathbf{P}\{\xi = k\} = p_k, \quad k = 0, 1, \ldots,$$

где $p_0 > 0$ и существует $k > 0$, для которого $p_k > 0$. В диссертации рассматривается подмножество $\mathfrak{F}_{N,n} \in \mathfrak{F}_N$ лесов с N деревьями и n некорневыми вершинами. Обозначим

$$F(z) = \sum_{k=0}^{\infty} p_k z^k.$$

Известно, что если уравнение $zF'(z) = F(z)$ имеет положительное решение, не превосходящее радиуса сходимости производящей функции $F(z)$, то лес Гальтона — Ватсона $\mathfrak{F}_{N,n}$ можно считать сгенерированным ветвящимся процессом с N начальными частицами, в котором распределение числа прямых потомков каждой частицы задается равенствами

$$p_k(\lambda) = \frac{\lambda^k p_k}{F(\lambda)}, \quad 0 < \lambda \leqslant 1, \quad k = 0, 1, \ldots, \tag{2}$$

при этом $\mathbf{E}\xi = 1$.

Частным случаем леса $\mathfrak{F}_{N,n}$ является множество корневых лесов с $N+n$ помеченными вершинами, на котором задано равномерное распределение вероятностей. В диссертации изучается предельное поведение числа цепей в таком случайном лесе.

Во второй главе рассматривается множество $S_{n,N}$ случайных подстановок степени n, имеющих N циклов. На множестве $S_{n,N}$ задается равномерное распределение вероятностей. Известно, что при изучении числовых характеристик таких случайных подстановок можно использовать обобщенную схему размещения (1), где η_i — случайная величина, равная длине i-го цикла случайно выбранной подстановки, $1 \leqslant i \leqslant N$, а независимые одинаково распределенные случайные величины $\xi_1, \ldots, \xi_N$ имеют распределение логарифмического ряда:

$$p_k = \mathbf{P}\{\xi_1 = k\} = \frac{\lambda^k}{-k \ln(1-\lambda)}, \ k = 1, 2, \ldots, \ 0 < \lambda < 1. \tag{3}$$

Такая обобщенная схема размещения может применяться также и при изучении объемов деревьев в случайном рекурсивном лесе $R_{n,N}$, поэтому все результаты о подстановках, полученные в диссертации, переносятся на рекурсивные леса.

Пусть μ_r — случайная величина, равная числу циклов длины r в случайно выбранной подстановке из $S_{n,N}$, распределение независимых одинаково распределенных случайных величин $\xi_1^{(r)}, \cdots, \xi_N^{(r)}$ задается равенствами

$$\mathbf{P}\left\{\xi_1^{(r)} = k\right\} = \mathbf{P}\{\xi_1 = k |\ \xi_1 \neq r\}, \ k = 0, 1, 2, \ldots$$

Введем обозначения:

$$\zeta_N = \xi_1 + \ldots + \xi_N,\ \zeta_N^{(r)} = \xi_1^{(r)} + \ldots + \xi_N^{(r)}.$$

Согласно хорошо известному свойству обобщенной схемы размещения, имеет место следующее утверждение.

Лемма 1.1.1. *Для любого* $k = 0, 1, \ldots, N$

$$\mathsf{P}\{\mu_r = k\} = C_N^k p_r^k (1 - p_r)^{N-k} \frac{\mathsf{P}\left\{\zeta_{N-k}^{(r)} = n - kr\right\}}{\mathsf{P}\{\zeta_N = n\}}.$$

При помощи этой леммы получение предельного распределения случайной величины μ_r при $n, N \to \infty$ сводится к доказательству локальных предельных теорем для сумм $\zeta_{N-k}^{(r)}$ и ζ_N независимых одинаково распределенных вспомогательных случайных величин и к оценке биномиальных вероятностей вида $C_N^k p_r^k (1 - p_r)^{N-k}$. При этом выбор параметра распределения (3) осуществляется наиболее удобным для доказательства способом. Во второй главе найдены предельные распределения случайной величины μ_r во всех зонах изменения параметров n, N, r. В доказанных теоремах параметр λ распределения (3) равен $\lambda = 1 - n^{-1}$ при $N/\ln n = O(1)$, а если $N/\ln n \to \infty$, то λ — решение уравнения

$$-\frac{\lambda}{(1-\lambda)\ln(1-\lambda)} = \frac{n}{N} \tag{4}$$

из интервала (0,1). Ниже приводятся лишь некоторые из полученных в диссертации результатов.

Теорема 2.1.5 *Пусть* $n, N \to \infty$ *так, что выполнено одно из условий:*

(1) $n/N \geqslant c > 1,\ N/\ln n \to \infty,\ Np_r \to \infty$;

(2) $n/N \to 1,\ (n-N)^2/N \to \infty,\ r = 1$;

(3) $n/N \to 1,\ (n-N)^3/N^2 \to \infty,\ r = 2$;

(4) $n/N \to 1,\ Np_r \to \infty,\ r \geqslant 3$.

Тогда

$$\mathsf{P}\{\mu_r = k\} = \frac{1}{\sigma_{rr}\sqrt{2\pi N}} e^{-u_r^2/2}\,(1 + o(1))$$

равномерно относительно k*, для которых* $u_r = (k - Np_r)/\sigma_{rr}\sqrt{N}$ *находится в любом фиксированном конечном интервале, где*

$$\sigma_{rr}^2 = p_r\left(1 - p_r - \frac{(m-r)^2 p_r}{\sigma^2}\right), \quad \sigma^2 = m\left(\frac{1}{1-\lambda} - m\right), \tag{5}$$

$$m = -\frac{\lambda}{(1-\lambda)\ln(1-\lambda)}. \tag{6}$$

Теорема 2.1.6 *Пусть выполнено одно из следующих условий:*

(1) $n, N, r \to \infty$ *так, что* $1 < c \leqslant n/N \leqslant c_1 < \infty$;

(2) $n, N \to \infty$ *так, что* $n/N \to \infty$, $N/\ln n \to \infty$;

(3) $n \to \infty$, $N = \gamma \ln n + o(\ln n)$, γ — *постоянная,* $0 < \gamma < \infty$;

(4) $n, N \to \infty$ *так, что* $n/N \to 1$, $n - N \to \infty$, $r \geqslant 3$.

Тогда

$$\mathsf{P}\{\mu_r = k\} = \frac{(Np_r)^k}{k!} e^{-Np_r}\,(1 + o(1))$$

равномерно относительно к, для которых $(k - Np_r)/\sqrt{Np_r}$ *находится в любом фиксированном конечном интервале.*

Пусть $[x]$ означает целую часть числа x.

Теорема 2.1.11 *Если* $N \geqslant 2$, $n \to \infty$, $N/\ln n \to 0$, $r = [\alpha n]$, $0 < \alpha = \alpha(n) \leqslant c < 1$, *и* $2r \neq n$ *при* $N = 2$, *то*

$$\mathsf{P}\{\mu_r = 0\} = 1 - (N-1)p_r \frac{e^\alpha}{1-\alpha} + o(Np_r),$$

$$\mathsf{P}\{\mu_r = 1\} = (N-1)p_r \frac{e^\alpha}{1-\alpha} + o(Np_r).$$

В диссертации доказано, что для любых г при $n \to \infty$, $N/\ln n \to 0$ выполняются соотношения $\mathbf{P}\{\mu_r = 0\} \to 1$, $\mathbf{P}\{\mu_r > 0\} \to 0$ (теоремы **2.1.11-2.1.14)**, то есть предельным распределением является вырожденное, а если рассмотреть скорость сходимости к предельному распределению, то асимптотическое поведение распределений μ_r можно описать с помощью закона Бернулли.

В третьей главе рассматриваются случайные подстановки из множества $S_{n,N}$ и леса Гальтона — Ватсона с N деревьями и п некорневыми вершинами.

Легко видеть, что если $n/N \to 1$, то асимптотически в среднем каждый цикл подстановки из $S_{n,N}$ состоит из одного элемента, и в этом случае можно считать все циклы компонентами малого объема. Из теорем 2.1.5, 2.1.6 следует, что при $r \geqslant 3$, $n/N \to 1$, $Np_r \to \infty$, где p_r определено в (3), предельное распределение случайной величины μ_r, равной числу циклов длины r в случайно выбранной подстановке, одновременно описывается нормальным законом и распределением Пуассона. Естественно возникает вопрос о том, какое приближение для распределения μ_r лучше.

Пусть

$$\beta = \frac{(m-r)^2 p_r}{\sigma^2},$$

где p_r, σ^2, m определены в (3), (5), (6). В главе 3 показано, что при $\beta = O\left(1/\sqrt{n-N}\right)$ лучшее приближение для распределения μ_r дает закон Пуассона; если $1/Np_r = o\left(\beta e^{-1/\beta}\right)$ и $1/\sqrt{n-N} = o(\beta)$, то более точным приближением для распределения μ_r является нормальный закон. При $1/\sqrt{n-N} = o(\beta)$, $\beta e^{-1/\beta} = O(1/Np_r)$ распределение Пуассона является лучшим, если $\lambda < \lambda_r$, где

$$\lambda_r = \left(\frac{c_2^2 r^3}{4c_1^2(1-r)^4 N}\right)^{\frac{1}{3r-6}} (1+o(1)),$$

$$c_1 = \sqrt{\frac{2}{\pi e}},\ c_2 = \frac{1+4e^{-3/2}}{3\sqrt{2\pi}},$$

и если $\lambda > \lambda_r$, то более точное приближение дает нормальный закон. В случае $\lambda = \lambda_r$ расстояния между распределением μ_r и предельными распределениями асимптотически равны.

Как отмечено выше, при $n \to \infty$, $N/\ln n \to 0$ предельным распределением случайной величины μ_r, равной числу циклов длины r в случайной подстановке из $S_{n,N}$, для любого r является вырожденное, то есть $\mathbf{P}\{\mu_r = 0\} \to 1$. Получается, что, хотя любая случайная подстановка имеет свою цикловую структуру, вероятность того, что μ_r примет какое-то ненулевое значение, при любом r стремится к нулю. Нам представляется, что понять причину этого явления в какой-то степени может помочь изучение поведения случайной величины $\eta_{(1)}$, равной минимальной длине цикла подстановки из $S_{n,N}$. Эта характеристика ранее не изучалась. В диссертации доказаны следующие теоремы.

Теорема 3.1.5. *Если $n, N \to \infty$ так, что $N/\ln n \to \infty$, то*

$$\mathbf{P}\left\{\eta_{(1)} = 1\right\} = 1 + o(1).$$

Теорема 3.1.6. *Если $n \to \infty$, $N = \gamma \ln n + o(\ln n)$, $0 < \gamma < \infty$, то для $r = 1, 2, \ldots$*

$$\mathbf{P}\left\{\eta_{(1)} \leqslant r\right\} = 1 - \exp\left\{-\gamma \sum_{k=1}^{r} \frac{1}{k}\right\} + o(1).$$

Теорема 3.1.7. *Пусть $n, N \to \infty$ так, что $n/N \to \infty$, $N/\ln n \to 0$, и пусть $r = n^{z/N}$, где z — положительная постоянная. Тогда*

$$\mathbf{P}\left\{\eta_{(1)} \leqslant r\right\} = 1 - e^{-z} + o(1).$$

Теорема 3.1.8. *Если $n \to \infty$, $r = n^{\alpha}$, $0 < \alpha < 1$, то для фиксированных α и $N \geqslant 2$*

$$\mathbf{P}\left\{\eta_{(1)} \leqslant r\right\} = 1 - (1 - \alpha)^{N-1} + o(1).$$

Согласно теореме 3.1.8, при $N = 2$ и фиксированных α, $0 < \alpha < 1$,

$$\mathsf{P}\left\{\eta_{(1)} \leqslant n^{\alpha}\right\} = \alpha + o(1),$$

то есть предельное распределение случайной величины $\log_n \eta_{(1)}$ является равномерным на интервале (0,1). Отсюда получаем, что в $S_{n,2}$ содержатся только такие подстановки, в которых циклы имеют длины вида n^{α} и $n - n^{\alpha}$, г $0 < \alpha < 1$, о при каждом фиксированном α вероятность наличия таких циклов стремится к нулю. Это объясняет, почему в этом случае $\mathsf{P}\{\mu_r > 0\} \to 0$ для любого r. Аналогичные рассуждения можно провести и при $N \geqslant 3$.

Также в третьей главе доказаны предельные теоремы для случайной величины $\eta^{(1)}$, равной минимальному объему дерева в лесе Гальтона — Ватсона $\mathfrak{F}_{N,n}$. Пусть j — наименьшее положительное целое число такое, что $p_j > 0$, а l — наименьшее натуральное, не кратное j, удовлетворяющее условию $p_{j+l} > 0$, если такого l нет, то положим $l = 0$. Пусть параметр λ распределения (2) определяется равенством

$$\frac{\lambda F'(\lambda)}{F(\lambda)} = \frac{n}{n+N}$$

при $n/N^2 \to 0$, а если $n/N^2 \geqslant C > 0$, то $\lambda = 1$. Пусть $\nu^{(1)}, \ldots, \nu^{(N)}$ — независимые одинаково распределенные случайные величины, равные объемам деревьев с корнями $1, \ldots, N$, соответственно, в лесе Гальтона — Ватсона $\mathfrak{F}_N$:

$$q_k = \mathsf{P}\left\{\nu^{(1)} = k+1\right\}, \quad k = 0, 1, \ldots$$

Теорема 3.1.9 *Пусть $n, N \to \infty$ так, что $N\lambda^{j+l} \to \infty$. Тогда*

$$\mathsf{P}\left\{\eta^{(1)} = 1\right\} = 1 + o(1).$$

Теорема 3.1.10 *Если $n \to \infty$, то для фиксированных $N \geqslant 2$ и $r = 1, 2, \ldots$*

$$\mathsf{P}\left\{\eta^{(1)} \leqslant r\right\} = 1 - \left(1 - \sum_{k=0}^{r-1} q_k\right)^{N-1} + o(1).$$

Теорема 3.1.9 показывает, что если $N \to \infty$, то, сколь велико бы ни было п, минимальный объем дерева асимптотически равен единице. Этот результат можно объяснить, по-видимому, тем, что если $n/N^2 \geqslant C > 0$, то, как известно, в лесе $\mathfrak{F}_{N,n}$ имеются деревья, объем которых имеет порядок п, а если $n/N^2 \to \infty$, то возникает так называемое гигантское дерево, содержащее п + $o(n)$ вершин.

В четвертой главе изучается предельное поведение числа пар в обобщенной схеме размещения, где под парой понимаются две различные частицы, попавшие в одну ячейку. Так, для подстановки, пару составляют два различных элемента подстановки, принадлежащих одному циклу, в корневом лесе с помеченными вершинами пара — две различные некорневые вершины, принадлежащие одному дереву. Обозначим через ν_1 число пар извлеченных шаров одинакового цвета в урновой схеме с разноцветными шарами, ν_2 — число пар в случайном корневом лесе с помеченными вершинами, ν_3 — число пар в случайной подстановке из $S_{n,N}$. Нетрудно видеть, что число различных цепей в случайном корневом лесе с помеченными вершинами равно $\tau = \nu_2 + n$, а число различных путей в графе случайной подстановки из $S_{n,N}$ равно $2\nu_3$. В четвертой главе доказаны следующие теоремы.

Теорема 4.1.1 *Пусть $m \geqslant 2$, $0 < n < mN$, $n, N \to \infty$ так, что $n^2/N \to \infty$, $n/mN \leqslant C < 1$. Тогда*

$$\mathbf{P}\{\nu_1 = k\} = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(k-a)^2}{2\sigma^2}\right\}(1 + o(1))$$

равномерно относительно целых неотрицательных k, для которых $(k-a)/\sigma$ находится в любом фиксированном конечном интервале, где

$$a = \frac{(m-1)n^2}{2mN}, \quad \sigma^2 = \frac{(m-1)(mN-n)^2 n^2}{2m^3N^3}.$$

Теорема 4.1.2 *Пусть $n, N \to \infty$ так, что $n/N \leqslant C < \infty$, $n^2/N \to \infty$. Тогда*

$$\mathbf{P}\{\tau - n = k\} = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{(k-a)^2}{2\sigma^2}\right\}(1 + o(1))$$

равномерно относительно целых неотрицательных k*, для которых* $(k-a)/\sigma$ *находится в любом фиксированном конечном интервале, где*

$$a = \frac{n^2(n+3N)}{2N^2}, \quad \sigma^2 = \frac{3n^2(n+N)^5}{2N^6}.$$

Теорема 4.1.3 *Пусть* $n, N \to \infty$ *так, что* $1 < c_1 \leqslant n/N \leqslant c_2 < \infty$. *Тогда*

$$\mathbf{P}\{\nu_3 = k\} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(k-a)^2}{2\sigma^2}\right\}(1+o(1))$$

равномерно относительно целых неотрицательных к, для которых $(k-a)/\sigma$ *находится в любом фиксированном конечном интервале, где*

$$a = \frac{n\lambda}{2(1-\lambda)}, \quad \sigma^2 = \frac{n\lambda}{4(1-\lambda)^3}\left(\frac{\lambda^2}{\ln(1-\lambda)+\lambda} + 2\right),$$

λ *— решение уравнения* (4) *из интервала* (0,1).

Пятая глава содержит некоторые статистические приложения. В частности, рассматривается следующая задача. Предположим, что из урны, содержащей K различных шаров, каждый из которых окрашен в один из N цветов, осуществляется равновероятный выбор без возвращения n шаров, $0 < n < K$. Статистическая гипотеза состоит в том, что число шаров каждого цвета одинаково и равно m, то есть $K = mN$. Для проверки этой гипотезы естественно использовать статистику

$$\chi^2 = \sum_{i=1}^{N} \frac{(\eta_i - n/N)^2}{n/N},$$

где η_i — число извлеченных шаров i-го цвета, $1 \leqslant i \leqslant N$.

Такую статистику можно использовать также для проверки статистических гипотез о равномерности распределения вершин по деревьям в случайном корневом лесе с помеченными вершинами и элементов подстановки из $S_{n,N}$ по циклам, г дη_i — случайная величина, равная числу некорневых вершин в i-ом дереве леса или, соответственно, длине i-ого цикла подстановки, $1 \leqslant i \leqslant N$. В случае, когда рассматриваемые гипотезы верны, асимптотики соответствующих статистик следуют из теорем **4.1.1-4.1.3** и легко проверяемого с помощью равенства $\eta_1 + \ldots + \eta_N = n$ соотношения

$$\mathsf{P}\left\{\chi^2 = \frac{2N}{n}x + N - n\right\} = \mathsf{P}\left\{\frac{1}{2}\sum_{i=1}^{N}\eta_i(\eta_i - 1) = x\right\},$$

справедливого для любого *x*.

Статистический критерий пустых ящиков хорошо известен. Обычно он применяется для проверки гипотезы о том, что выборка, состоящая из *n* независимых наблюдений, взята из непрерывного распределения *G(x)*. Критерий пустых ящиков основывается на классической задаче о размещении *n* дробинок по *N* ящикам, при этом N интервалам, на которые делится числовая ось, соответствуют ящики, а наблюдениям — размещаемые по ящикам дробинки. Критерий строится на основе статистики μ_0, равной числу пустых ящиков при равновероятном размещении *n* дробинок по *N* ящикам. Следуя этим идеям и используя современные вероятностные методы дискретной математики, можно предложить аналогичные критерии для статистического анализа некоторых комбинаторных объектов. В разделах 5.3, 5.4 диссертации показана применимость критерия пустых ящиков для проверки статистической гипотезы о равномерности распределения вероятностей на множестве лесов Гальтона — Ватсона $\mathfrak{F}_{N,n}$.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи

1. Павлов Ю. Л., Черепанова Е. В. Предельные распределения числа пар в обобщенной схеме размещения. // Дискретная математика, 2002, **14,** №3, с.149-159.

2. Черепанова Е. В. Предельные распределения числа циклов заданной длины в случайной подстановке с известным числом циклов. // Дискретная математика, 2003, 15, №3, с.128-144.

3. Черепанова Е. В. Асимптотика статистики x^2 *в* некоторых комбинаторных задачах. // Тр. Института прикл. матем. исслед. Карельского НЦ РАН, 2003, 4, с.129-135.

4. Черепанова Е. В. Критерий пустых ящиков для случайных лесов. // Тр. Института прикл. матем. исслед. Карельского НЦ РАН, 2003, 4, с.136-143.

Тезисы докладов

5. Павлов Ю. Л., Черепанова Е. В. О числе путей в случайной подстановке с известным числом циклов. // Обозрение прикладной и промышленной математики, 2002, 9, в.2, с.431-432.

6. Черепанова Е. В. Распределение числа пар объектов в одной урновой схеме. // Тез. докл. науч. конф. "Карелия и РФФИ", Петрозаводск, 2002, с.99-100.

7. Cherepanova E. V. Limit distribution of number of chains in a random forest. // Information processes, 2002, 2, №2, pp,165.

8. Черепанова Е. В. Одна задача о случайных подстановках с известным числом циклов. // Обозрение прикладной и промышленной математики, 2003,10, в.1, с.251-252.

9. Павлов Ю. Л., Черепанова Е. В. Предельные распределения минимальной длины цикла в случайной подстановке с известным числом циклов. // Обозрение прикладной и промышленной математики, 2003, 10, в.2, с.357-358.

Изд. лиц. № 00041 от 30.08.99. Подписано в печать 22.03.04. Формат 60x84 1/16.
Бумага офсетная. Гарнитура «Times». Печать офсетная.
Уч.-изд. л. 1,0. Усл. печ. л. 1,0. Тираж 100 экз. Изд. № 34. Заказ № 421

Карельский научный центр РАН
185003, Петрозаводск, пр. А. Невского, 50
Редакционно-издательский отдел

3

№ 11775

471