G. I. Ivchenko, Yu. I. Medvedev, Random combinatorial objects in a general parametric model,
Tr. Diskr. Mat., 2007, Volume 10, 73–86

Труды по дискретной математике
Том 10, с. 73–86
М.: ФИЗМАТЛИТ, 2007

СЛУЧАЙНЫЕ КОМБИНАТОРНЫЕ ОБЪЕКТЫ В ОБЩЕЙ ПАРАМЕТРИЧЕСКОЙ МОДЕЛИ

Г. И. ИВЧЕНКО, Ю. И. МЕДВЕДЕВ

Предлагается новый подход к проблемам рандомизации и классификации в вероятностной комбинаторике, связанный с введением на произвольном множестве разложимых комбинаторных объектов некоторой общей параметрической меры, обладающей достаточным числом степеней свободы, чтобы удовлетворить потребности криптографической практики в рассмотрении неравновероятных комбинаторных объектов самой различной природы.

§ 1. ВВЕДЕНИЕ

Мы рассматриваем комбинаторные объекты, разложимые на отдельные компоненты (далее — разложимые комбинаторные объекты). К таковым относятся:
подстановки, т. е. взаимно-однозначные отображения конечного множества $X_n = \{1, 2, \ldots, n\}$ в себя, для которых компонентами являются циклы;
однозначные отображения множества X_n в себя, разложимые на компоненты связности;
разбиения множества X_n на блоки и целых чисел на слагаемые;
многочлены над конечным полем $\mathbf{GF}(q)$ (q — простое или степень простого числа), разложимые на неприводимые множители, и т. д.

Все эти объекты имеют непосредственные криптографические приложения в различных направлениях.

Пусть n — характеристический параметр рассматриваемых объектов, их вес (для отображений и разбиений конечного множества n — число элементов, для графов n — число вершин, для многочленов n — степень и т. д.).

Символом $\mathcal{K}_n$ обозначим множество разложимых объектов веса n, и для конкретного объекта $K \in \mathcal{K}_n$ пусть $c_i(n)$ — число его компонент размера (веса) i ($i = 1, 2, \ldots, n$). Набор $c(n) = (c_1(n), c_2(n), \ldots, c_n(n))$ называется структурой объекта K. Так как $ic_i(n)$ — вес всех компонент размера i объекта K, то

$$\sum_{i=1}^{n} ic_i(n) = n.$$

Наконец, пусть $N(n, c(n))$ обозначает число объектов из $\mathcal{K}_n$ со структурой $c(n)$.

Основной интерес для криптографических приложений представляет исследование различных структурных свойств тех или иных комбинаторных объектов.

Наиболее эффективным методом решения подобных задач является теоретико-вероятностный подход, когда на множестве $\mathcal{K}_n$ вводится та или иная вероятностная мера $\mathbf{P}$, приписывающая каждому объекту $K \in \mathcal{K}_n$ вероятность $\mathbf{P}(K)$ его наблюдения. Так возникают случайные комбинаторные объекты: случайные подстановки, случайные отображения, случайные разбиения, случайные многочлены и т. д. При этом структура $c(n)$ случайного объекта становится случайной величиной, для изучения свойств которой применяются различные вероятностные методы и особенно эффективно — предельные теоремы теории вероятностей, позволяющие исследовать асимптотические особенности изучаемых объектов для больших значений параметра n (типичная ситуация и наиболее актуальная проблематика для криптографической практики).

До последнего времени большинство исследований в этой области ограничивалось рассмотрением лишь равновероятных объектов, когда $\mathbf{P}$ — равномерная мера на множестве $\mathcal{K}_n$ либо на некотором выделенном его подмножестве. В этом важнейшем случае все рассматриваемые задачи, несмотря на вероятностную терминологию, по существу представляют собой перечисленные задачи комбинаторного анализа. Однако для современных приложений все более актуальной становится задача, требующая отказа от такого существенного ограничения, т. е. необходимости рассмотрения также и различного типа неравновероятных объектов. Это неоднократно отмечалось специалистами-криптографами и в научной литературе, в этом направлении имеются отдельные результаты, но общие подходы до сих пор отсутствовали.

В настоящем исследовании предлагается такой общий подход к проблеме рандомизации, в основе которого лежит введение на множестве $\mathcal{K}_n$ некоторой общей параметрической меры с большим числом степеней свободы, позволяющей изучать вероятностно-статистические свойства случайных комбинаторных объектов в самых различных ситуациях, встречающихся в криптографической практике, включая неравновероятные схемы различных типов. Идея и некоторые элементы этого подхода аннотированы в [1–3].

§ 2. ОПИСАНИЕ МОДЕЛИ

Пусть $\mathcal{K}_n$ — произвольное множество разложимых комбинаторных объектов веса n. Зададим на этом множестве вероятностную меру, зависящую от параметра $\theta = (\theta_1, \theta_2, \ldots, \theta_n)$, $\theta_i \geqslant 0$, в соответствии с которой произвольный

объект $K \in \mathcal{K}_n$ со структурой $c(n) = a = (a_1, a_2, \ldots, a_n)$ наблюдается с вероятностью, пропорциональной $\prod_i \theta_i^{a_i}$, а именно:

$$\mathbf{P}_\theta(K) = I\left(\sum_{i=1}^{n} i a_i = n\right) \frac{\prod_{i=1}^{n} \theta_i^{a_i}}{H_n(\theta)}, \tag{1}$$

где $I(\cdot)$ — индикатор и $H_n(\theta)$ — необходимый нормирующий множитель, который мы будем называть статистической суммой множества $\mathcal{K}_n$. Из условия нормировки

$$1 = \sum_{K \in \mathcal{K}_n} \mathbf{P}_\theta(K) = \sum_{a:\, \sum i a_i = n} N(n, a) \prod_{i=1}^{n} \theta_i^{a_i} \Big/ H_n(\theta)$$

следует, что

$$H_n(\theta) = \sum_{a:\, \sum i a_i = n} N(n, a) \prod_{i=1}^{n} \theta_i^{a_i}. \tag{2}$$

Соотношениями (1)–(2) и определяется общая параметрическая модель случайного объекта из множества $\mathcal{K}_n$. Достоинством этой модели является наличие в ней большого числа степеней свободы: зависимость от n-мерного «управляемого» параметра $\theta = (\theta_1, \ldots, \theta_n)$ придает ей большую гибкость, так как, выбирая соответствующим образом значения этого параметра, мы можем задавать на множестве $\mathcal{K}_n$ или его подмножествах различные распределения вероятностей, отвечающие наиболее точно интересующим нас характеристикам реальных комбинаторных объектов. Такой подход открывает, на наш взгляд, новые возможности в вероятностной комбинаторике, позволяет решать не только стандартные перечислительные задачи (подсчет числа объектов с заданными структурными свойствами), но и проводить более глубокий вероятностный и статистический анализ случайных комбинаторных объектов с произвольной, в том числе неравновероятностной метрикой, и тем самым более полно отвечать на запросы современной практики.

Проиллюстрируем сказанное несколькими примерами.

1. Если в (1) и (2) положить $\theta = (1, 1, \ldots, 1)$, то получим классическую равновероятную модель, когда каждый объект $K \in \mathcal{K}_n$ наблюдается (выбирается) с одной и той же вероятностью $|\mathcal{K}_n|^{-1}$, где

$$|\mathcal{K}_n| = H_n(1) = \sum_{a:\, \sum i a_i = n} N(n, a)$$

есть общее число объектов (объем) множества $\mathcal{K}_n$.

Большинство результатов вероятностной комбинаторики получено в рамках именно этой модели (В. Л. Гончаров, В. Е. Степанов, В. Н. Сачков, В. Ф. Колчин, П. Эрдеш, А. Реньи и др.).

2. Для заданного подмножества $R \subseteq X_n$ обозначим через $\mathcal{K}_n(R)$ подмножество тех объектов $K \in \mathcal{K}_n$, веса компонент которых принадлежат R. Полагая в (1) $\theta_i = 0$ для $i \notin R$, получим меру, сосредоточенную на $\mathcal{K}_n(R)$. Тем самым мы имеем возможность изучать различные структурные свойства комбинаторных объектов с учетом произвольного ограничения R. Типичными примерами задания R являются: $R = \{i\colon i \leqslant s\}$ при некотором $s < n$ (веса компонент не превосходят s) либо $R = \{i\colon i \equiv m \,(\mathrm{mod}\, d)\}$ при некоторых натуральных m и d и т. д. Если дополнительно $\theta_i = 1$ для $i \in R$, то получаем равновероятную модель R-объектов, которая и рассматривалась ранее в литературе (см. [4–6] и библиографию в них).

3. Если в (1) положить $\theta_1 = \ldots = \theta_n = \theta > 0$, то получим однопараметрическую меру, приписывающую объектам $K \in \mathcal{K}_n$ вероятности, пропорциональные $\theta^{|K|}$, где $|K| = c_1(n) + \ldots + c_n(n)$ — общее число компонент объекта K. Впервые такая неравновероятная мера рассматривалась Эвансом [7] для случайных подстановок, т. е. когда $\mathcal{K}_n$ есть множество $\mathcal{S}_n$ всех n-подстановок. Для этого случая, как известно, число подстановок циклового класса $a = (a_1, a_2, \ldots, a_n)$ есть

$$N(n, a) = I\left(\sum_{i=1}^{n} i a_i = n\right) \frac{n!}{\prod_i a_i! i^{a_i}},$$

и формула (2) принимает вид

$$H_n(\theta) = n! \sum_{a\colon \sum i a_i = n} \prod_{i=1}^{n} \left(\frac{\theta_i}{i}\right)^{a_i} \frac{1}{a_i!} = n! \,[z^n] \exp\left\{\sum_{i=1}^{n} \frac{z^i}{i} \theta_i\right\},$$

где

$$[z^n] f(z) = \mathrm{coef}_{z^n} f(z).$$

В модели Эванса (при $\theta_1 = \ldots = \theta_n = \theta > 0$)

$$H_n(\theta) = n! \,[z^n] \exp\{-\theta \ln(1 - z)\} = n! \,[z^n] (1 - z)^{-\theta} =$$

$$= n! (-1)^n \binom{-\theta}{n} = \theta(\theta + 1) \ldots (\theta + n - 1) \equiv \theta_{(n)}$$

и формула (1) принимает вид

$$\mathbf{P}_\theta(K) = I\left(\sum_{i=1}^{n} i a_i = n\right) \frac{\theta^{|K|}}{\theta_{(n)}}, \quad K \in \mathcal{S}_n.$$

Модель Эванса и ее аналоги для других комбинаторных объектов рассматривались многими авторами (см. [8–11] и библиографию в них).

§ 3. РАСПРЕДЕЛЕНИЕ ЦИКЛОВОЙ СТРУКТУРЫ

В модели (1) структура $c(n) = (c_1(n), c_2(n), \ldots, c_n(n))$ случайного объекта имеет распределение

$$\mathbf{P}_\theta(c(n) = a) = I\left(\sum_{i=1}^{n} i a_i = n\right) N(n, a) \frac{\prod_{i=1}^{n} \theta_i^{a_i}}{H_n(\theta)}. \tag{3}$$

Отсюда с учетом (2) легко выписать ее производящую функцию:

$$F_{n,\theta}(t) = \mathbf{E}_\theta \prod_{i=1}^{n} t_i^{c_i(n)} = \sum_{a:\, \sum i a_i = n} \mathbf{P}_\theta(c(n) = a) \prod_{i=1}^{n} t_i^{a_i} =$$

$$= \sum_{a:\, \sum i a_i = n} N(n, a) \prod_{i=1}^{n} (t_i \theta_i)^{a_i} \Big/ H_n(\theta) = \frac{H_n(t \bullet \theta)}{H_n(\theta)}, \tag{4}$$

где $t \bullet \theta = (t_1\theta_1, t_2\theta_2, \ldots, t_n\theta_n)$.

Соотношение (4) может быть основой для изучения различных структурных свойств комбинаторных объектов в рамках общей модели (1).

Эффективность использования этого соотношения зависит от сложности явного выражения статистической суммы $H_n(\theta)$, определенной в (2). Для большинства известных комбинаторных объектов функция $H_n(\theta)$ может быть выписана в явной форме, имеющей один из трех возможных типов, что одновременно даст естественную классификацию комбинаторных объектов в терминах этой функции.

Важнейшими характеристиками, определяющими вид статистической суммы $H_n(\theta)$, являются числа m_i соответствующих объектов веса i, состоящих из единственной компоненты. Эти числа для многих объектов в комбинаторике хорошо известны.

Например: для подстановок

$$m_i = (i-1)!;$$

для однозначных отображений

$$m_i = (i-1)! \sum_{j=0}^{i-1} \frac{i^j}{j!};$$

для нормированных (т. е. со старшим коэффициентом 1) многочленов над конечным полем $\mathbf{GF}(q)$

$$m_i = \frac{1}{i} \sum_{r|i} \mu(r) q^{i/r},$$

где $\mu(r)$ — функция Мебиуса: $\mu(1) = 1$, $\mu(r) = 0$, если r делится на квадрат простого числа, в противном случае $\mu(r) = (-1)^k$, где k — количество простых

множителей числа r (эта формула для m_i достаточно сложна, но при вычислении этих чисел можно использовать приближенное равенство $im_i \approx q^i$, или двустороннее неравенство $q^i - q^{i/2-1} \leqslant im_i \leqslant q^i$);
для разбиений конечных множеств и целых чисел $m_i = 1\ \forall i$.

Через числа m_i во многих случаях можно выписать явные формулы для подсчета числа $N(n, a)$ комбинаторных объектов с заданной структурой $a = (a_1, a_2, \ldots, a_n)$. Известны три общих типа комбинаторных объектов, для обозначения которых в литературе используются термины ансамбли, мультимножества и селекции. К типу ансамбля относятся, в частности, подстановки, однозначные отображения конечных множеств, разбиения конечных множеств, графы. Для них числа $N(n, a)$ и соответствующая статистическая сумма $H_n(\theta)$ имеют следующие представления:

$$N(n, a) = I\Bigl(\sum_{i=1}^{n} ia_i = n\Bigr) n! \prod_{i=1}^{n} \Bigl(\frac{m_i}{i!}\Bigr)^{a_i} \frac{1}{a_i!} \tag{5}$$

и (см. (2))

$$H_n(\theta) = n! \sum_{a:\, \sum ia_i = n} \prod_{i=1}^{n} \Bigl(\frac{m_i\theta_i}{i!}\Bigr)^{a_i} \frac{1}{a_i!} = n!\,[z^n] \exp\Bigl\{\sum_{i=1}^{n} \frac{z^i}{i!}\, m_i\theta_i\Bigr\}. \tag{6}$$

Таким образом, для ансамблей статистическая сумма имеет экспоненциальный тип.

К типу мультимножества (иначе, неупорядоченные выборки с повторениями элементов) относятся такие классы, как нормированные многочлены над конечным полем и разбиения целых чисел на слагаемые. Для них

$$N(n, a) = I\Bigl(\sum_{i=1}^{n} ia_i = n\Bigr) \prod_{i=1}^{n} \binom{m_i + a_i - 1}{a_i} \tag{7}$$

и

$$H_n(\theta) = \sum_{a:\, \sum ia_i = n} \prod_{i=1}^{n} \binom{m_i + a_i - 1}{a_i} \theta_i^{a_i} = [z^n] \prod_{i=1}^{n} \bigl(1 - z^i\theta_i\bigr)^{-m_i}. \tag{8}$$

Для таких случаев будем говорить, что функция $H_n(\theta)$ имеет отрицательный биномиальный тип.

Наконец, к типу селекций (иначе, неупорядоченных выборок без повторения элементов) относятся разбиения целых чисел на различные слагаемые и свободные от квадратов многочлены над полем $\mathbf{GF}(q)$. Для них числа $N(n, a)$ имеют вид

$$N(n, a) = I\Bigl(\sum_{i=1}^{n} ia_i = n\Bigr) \prod_{i=1}^{n} \binom{m_i}{a_i}, \tag{9}$$

а статистическая сумма $H_n(\theta)$ имеет биномиальный тип:

$$H_n(\theta) = \sum_{a:\,\sum ia_i = n} \prod_{i=1}^{n} \binom{m_i}{a_i} \theta_i^{a_i} = [z^n] \prod_{i=1}^{n} (1 + z^i \theta_i)^{m_i}. \tag{10}$$

Таким образом, каждый класс объектов во всех трех типах структур определяется характеристикой $m = (m_1, \ldots, m_n)$. Поэтому статистическую сумму $H_n(\theta)$ для конкретного класса с характеристикой m мы будем иногда обозначать $H_{n,m}(\theta)$.

В итоге можно констатировать, что для широкого круга комбинаторных объектов соотношения (4), (6), (8) и (10) дают общий математический аппарат исследования их различных вероятностно-статистических свойств в рамках общей модели (1).

Аналогичную методологию, в принципе, можно применять для изучения и иных, более сложных, комбинаторных конструкций, для которых удается получить явное представление для соответствующей статистической суммы.

В качестве иллюстрации использования соотношения (4) получим общую формулу для смешанных факториальных моментов случайной структуры $c(n)$ в случае ансамблей. Используя обозначение $(x)_r = x(x-1)\ldots(x-r+1)$, $r \geqslant 1$, $(x)_0 = 1$, для $m = \sum_i ir_i \leqslant n$ из (4) и (6) получаем, что

$$\mathbf{E}_\theta \prod_i (c_i)_{r_i} = \frac{\partial^{\sum_i r_i} F_{n,\theta}(t)}{\prod_i \partial t_i^{r_i}}\bigg|_{t=1} = \frac{1}{H_n(\theta)} \frac{\partial^{\sum_i r_i} H_n(t \bullet \theta)}{\prod_i \partial t_i^{r_i}}\bigg|_{t=1} =$$

$$= \frac{n!}{H_n(\theta)} \prod_i \left(\frac{m_i\theta_i}{i!}\right)^{r_i} [z^{n-m}] \exp\left\{\sum_{i=1}^{n-m} \frac{z^i}{i!}\, m_i\theta_i\right\} =$$

$$= \frac{n!}{(n-m)!} \frac{H_{n-m}(\theta)}{H_n(\theta)} \prod_i \left(\frac{m_i\theta_i}{i!}\right)^{r_i}. \tag{11}$$

В частности, формулы для средних значений имеют вид

$$\mathbf{E}_\theta c_i(n) = \frac{n!}{(n-i)!} \frac{m_i\theta_i}{i!} \frac{H_{n-i}(\theta)}{H_n(\theta)}. \tag{12}$$

§ 4. О МЕТОДАХ АНАЛИЗА МОДЕЛИ

Анализ модели (1) может быть основан на классических методах производящих и характеристических функций в сочетании с методом перевала (в асимптотических подходах), методе моментов, методе условных распределений, приводящем в данном случае к независимым, но разнораспределенным случайным величинам в схеме серий, наконец, методе рандомизации основ-

ного параметра n, также приводящем к независимым случайным величинам. Опишем детально два последних.

1. *Метод условных распределений.* Введем независимые в совокупности случайные величины (с. в.) $Z_1, Z_2, \ldots Z_n$, которые для ситуаций экспоненциального типа (6) имеют распределение Пуассона:

$$L_\theta(Z_i) = \Pi(\lambda_i), \quad \lambda_i = \frac{\theta_i m_i}{i!}, \quad i = 1, 2, \ldots, n;$$

для ситуаций, описываемых соотношением (8), — отрицательное биномиальное распределение:

$$L_\theta(Z_i) = \overline{\mathrm{Bi}}(m_i, \theta_i), \quad \theta_i \in (0,1), \quad i = 1, 2, \ldots, n;$$

наконец, для биномиального типа (10) Z_i имеет биномиальное распределение $\mathrm{Bi}(m_i, p_i)$ с $p_i = \dfrac{\theta_i}{1+\theta_i}$, $i = 1, 2, \ldots, n$.

Теорема 1. *Распределение случайной структуры*

$$c(n) = (c_1(n), c_2(n), \ldots, c_n(n))$$

в модели (1) *имеет представление в виде следующего условного распределения*:

$$L_\theta(c(n)) = L_\theta(Z_1, Z_2, \ldots, Z_n \mid T_n = n), \tag{13}$$

где $T_n = \sum_{i=1}^n iZ_i$.

Доказательство для экспоненциального случая следует из цепочки соотношений:

$$\mathbf{P}_\theta(Z_i = a_i, i = 1, \ldots, n \mid T_n = n) = \frac{\mathbf{P}_\theta(Z_i = a_i, i = 1, \ldots, n, T_n = n)}{\mathbf{P}_\theta(T_n = n)} =$$

$$= I\Big(\sum_{i=1}^n ia_i = n\Big) \prod_{i=1}^n \mathbf{P}_\theta(Z_i = a_i) \Big/ \sum_{b:\, \sum ib_i = n} \prod_{i=1}^n \mathbf{P}_\theta(Z_i = b_i) =$$

$$= e^{-\sum \lambda_i} I\Big(\sum_{i=1}^n ia_i = n\Big) \prod_{i=1}^n \frac{\lambda_i^{a_i}}{a_i!} \Big/ e^{-\sum \lambda_i} \sum_{b:\, \sum ib_i = n} \prod_{i=1}^n \frac{\lambda_i^{b_i}}{b_i!} =$$

$$= \frac{1}{n!} H_n(\theta) \mathbf{P}_\theta(c(n) = a) \Big/ \sum_{b:\, \sum ib_i = n} \frac{H_n(\theta)}{n!} \mathbf{P}_\theta(c(n) = b) = \mathbf{P}_\theta(c(n) = a),$$

поскольку

$$\sum_{b:\, \sum ib_i = n} \mathbf{P}_\theta(c(n) = b) = 1.$$

Для остальных двух случаев рассуждения аналогичны.

Замечание. Из предыдущих соотношений следует также, что для объектов экспоненциального типа (см. (6))

$$\mathbf{P}_\theta(T_n = n) = e^{-\sum \lambda_i} \frac{H_n(\theta)}{n!} = \exp\left\{-\sum_{i=1}^{n} \theta_i \frac{m_i}{i!}\right\} [z^n] \exp\left\{\sum_{i=1}^{n} \frac{z^i}{i!} m_i \theta_i\right\};$$

для объектов отрицательного биномиального типа (см. (8))

$$\mathbf{P}_\theta\,(T_n = n) = H_n(\theta) \prod_{i=1}^{n} (1-\theta_i)^{m_i} = \prod_{i=1}^{n} (1-\theta_i)^{m_i} [z^n] \prod_{i=1}^{n} \left(1 - z^i \theta_i\right)^{-m_i};$$

для объектов биномиального типа (см. (10))

$$\mathbf{P}_\theta\,(T_n = n) = H_n(\theta) \prod_{i=1}^{n} (1+\theta_i)^{-m_i} = \prod_{i=1}^{n} (1+\theta_i)^{-m_i} [z^n] \prod_{i=1}^{n} \left(1 + z^i \theta_i\right)^{m_i}.$$

2. *Метод рандомизации параметра* n. Пусть параметр n является реализацией неотрицательной целочисленной с. в. η, имеющей распределение типа степенного ряда:

$$\mathbf{P}_\theta(\eta = n) = f_\theta^{-1}(x) x^n [x^n] f_\theta(x), \quad n = 0, 1, 2, \ldots, \tag{14}$$

где для объектов экспоненциального типа (см. (6)) функция $f_\theta(x)$ имеет вид

$$f_\theta(x) = \exp\left\{\sum_{i=1}^{\infty} \lambda_i x^i\right\}, \quad \lambda_i = \frac{\theta_i m_i}{i!}, \tag{15}$$

для объектов отрицательного биномиального типа (см. (8))

$$f_\theta(x) = \prod_{i=1}^{\infty} \left(1 - \theta_i x^i\right)^{-m_i}, \tag{16}$$

для объектов биномиального типа (см. (10))

$$f_\theta(x) = \prod_{i=1}^{\infty} \left(1 + \theta_i x^i\right)^{m_i} \tag{17}$$

(здесь $x > 0$ — «свободный» параметр, значение которого может быть выбрано произвольно из области аналитичности соответствующей функции $f_\theta(x)$).

Теорема 2. *При указанной рандомизации параметра n элементы случайной структуры $c_i(\eta)$, $i \geqslant 1$, будут независимыми с. в.; при этом для объектов экспоненциального типа*

$$L_\theta(c_i(\eta)) = \Pi(\lambda_i x^i),$$

для объектов отрицательного биномиального типа

$$L_\theta(c_i(\eta)) = \overline{\mathrm{Bi}}\,(m_i, \theta_i x^i),$$

для объектов биномиального типа

$$L_\theta(c_i(\eta)) = \mathrm{Bi}\,(m_i, p_i(x)), \quad p_i(x) = \frac{\theta_i x^i}{1+\theta_i x^i}.$$

Доказательство проведем для экспоненциального случая (для двух других случаев рассуждения аналогичны).

По формуле полной вероятности из (14)–(15) с учетом (3), (5) и (6) имеем

$$\mathbf{E}_\theta \prod_{i\geqslant 1} t_i^{c_i(\eta)} = \sum_{n\geqslant 0} \mathbf{P}_\theta(\eta = n)\mathbf{E}_\theta \prod_{i=1}^{n} t_i^{c_i(n)} =$$

$$= \sum_{n\geqslant 0} \frac{H_n(\theta)}{n! f_\theta(x)}\, x^n \sum_{a:\,\sum i a_i = n} \frac{n!}{H_n(\theta)} \prod_{i=1}^{n} \frac{(t_i\lambda_i)^{a_i}}{a_i!} =$$

$$= f_\theta^{-1}(x) \sum_{n\geqslant 0} \sum_{a:\,\sum i a_i = n} \prod_{i=1}^{n} \frac{(t_i\lambda_i x^i)^{a_i}}{a_i!} =$$

$$= f_\theta^{-1}(x) \prod_{i\geqslant 1} \sum_{a_i=0}^{\infty} \frac{(t_i\lambda_i x^i)^{a_i}}{a_i!} = \prod_{i\geqslant 1} \exp\left\{\lambda_i x^i (t_i - 1)\right\},$$

что эквивалентно первому утверждению теоремы.

Замечание. Параметр рандомизации x можно подбирать специальным образом в зависимости от цели исследования. Примером реализации изложенного метода является работа [12], в которой проведено детальное исследование структуры подстановки случайной степени в рамках модели Эванса.

§ 5. СЛУЧАЙНЫЕ КОМБИНАТОРНЫЕ ОБЪЕКТЫ ТИПА АНСАМБЛЯ

1. Рассмотрим сначала в качестве примера применения изложенной выше методологии наиболее известный комбинаторный объект — случайные n-подстановки, т. е. взаимно-однозначные отображения множества $X_n = \{1, 2, \ldots, n\}$ в себя. К уже сказанному в § 2 (пример 3) добавим, что, как отмечено в § 3, подстановки относятся к объектам экспоненциального типа с характеристиками $m_i = (i-1)!$ (см. (6)), поэтому производящая функция для цикловой структуры случайных R-подстановок для произвольного подмножества $R \subseteq X_n$ имеет в модели (1) вид (см. (4))

$$F_{n,\theta}(t; R) = \mathbf{E}_\theta \prod_{i\in R} t_i^{c_i(n)} = [z^n] \exp\left\{\sum_{i\in R} \frac{z^i}{i}\, t_i\theta_i\right\} \Big/ [z^n] \exp\left\{\sum_{i\in R} \frac{z^i}{i}\, \theta_i\right\}. \qquad (18)$$

Формула же (11) для факториальных моментов в этом случае конкретизируется и приобретает вид

$$\mathbf{E}_\theta \prod_{j\in R} (c_j)_{r_j} = \prod_{j\in R} \left(\frac{\theta_j}{j}\right)^{r_j} [z^{n-m}] \exp\left\{\sum_{i\in R} \frac{z^i}{i}\,\theta_i\right\} \Big/ [z^n] \exp\left\{\sum_{i\in R} \frac{z^i}{i}\,\theta_i\right\}, \quad (19)$$

где $m = \sum_{j\in R} jr_j \leqslant n$.

Соотношения (18)–(19) открывают новые возможности для изучения случайных подстановок. Их детальное исследование как в общей модели (1), так и в ее различных частных случаях (представляющих интерес для приложений) ждет своего решения. Здесь мы ограничимся рассмотрением одного конкретного иллюстративного примера, а именно, пусть $\theta_i = \theta/(i-1)!$, $i \geqslant 1$, $\theta > 0$, $R = X_n$, и будем в этой ситуации интересоваться распределением общего числа циклов $c(n) = c_1(n) + \ldots + c_n(n)$ случайной n-подстановки.

Производящая функция величины $c(n)$ получается из (18) при $t_i = t$, $i \geqslant 1$, и имеет вид

$$F_{n,\theta}(t) = \mathbf{E}_\theta t^{c(n)} = [z^n] \exp\{t\theta(e^z - 1)\} / [z^n] \exp\{\theta(e^z - 1)\} =$$

$$= \sum_{k=1}^{n} \sigma(n,k)\,(t\theta)^k \Big/ \sum_{k=1}^{n} \sigma(n,k)\,\theta^k, \quad (20)$$

где $\sigma(n,k)$ — числа Стирлинга второго рода (см. [6, с. 246]).

Таким образом, распределение числа циклов случайной подстановки в данной модели имеет вид

$$\mathbf{P}_\theta(c(n) = r) = \frac{\sigma(n,r)\,\theta^r}{\sum_{k=1}^{n} \sigma(n,k)\,\theta^k}, \quad r = 1, 2, \ldots, n. \quad (21)$$

2. Из представления (6) следует, что такое же распределение имеет и общее число блоков случайного разбиения множества X_n в модели, когда каждому разбиению K приписывается вес, пропорциональный $\theta^{|K|}$. Тем самым соотношения (20)–(21) могут быть основой и для изучения неравновероятных разбиений конечных множеств (до сих пор в литературе рассматривались лишь равновероятные разбиения, соответствующие случаю $\theta = 1$ [6]).

Данная связь между классами ансамбля — разбиениями конечного множества и подстановками — неслучайна. На самом деле имеет место более общее утверждение.

ТЕОРЕМА «переноса». *Пусть $\mathcal{K}_n$ и $\mathcal{K}'_n$ — два различных класса ансамбля. Пусть $K \in \mathcal{K}_n$ и $K' \in \mathcal{K}'_n$ — произвольные объекты этих классов с соответствующими характеристиками $m = (m_1, \ldots, m_n)$, $m' = (m'_1, \ldots, m'_n)$ и параметрами $\theta = (\theta_1, \ldots, \theta_n)$ и $\theta' = (\theta'_1, \ldots, \theta'_n)$. Тогда, если*

$$\theta'_i = \frac{m_i}{m'_i}\,\theta_i, \quad i = 1, \ldots, n,$$

то для любой заданной структуры $c(n) = a = (a_1, \ldots, a_n)$ *имеет место равенство*

$$\mathbf{P}_\theta(K) = I\Big(\sum_{i=1}^n i a_i = n\Big)\frac{\prod_{i=1}^n \theta_i^{a_i}}{H_{n,m}(\theta)} = I\Big(\sum_{i=1}^n i a_i = n\Big)\frac{\prod_{i=1}^n \theta_i'^{a_i}}{H_{n,m'}(\theta')} = \mathbf{P}_{\theta'}(K').$$

Доказательство. Из формулы (6) следует, что

$$H_{n,m}(\theta) = H_{n,m'}(\theta'),$$

а из формул (3) и (5) получаем

$$\mathbf{P}_{m,\theta}(c(n) = a) = I\Big(\sum_{i=1}^n i a_i = n\Big)\frac{n!}{H_{n,m}(\theta)}\prod_{i=1}^n \frac{(m_i\theta_i)^{a_i}}{i!^{a_i}a_i!} =$$

$$= I\Big(\sum_{i=1}^n i a_i = n\Big)\frac{n!}{H_{n,m'}(\theta')}\prod_{i=1}^n \frac{(m_i'\theta_i')^{a_i}}{i!^{a_i}a_i!} = \mathbf{P}_{m',\theta'}(c(n) = a).$$

Эта простая теорема для классов объектов ансамбля дает возможность перенести любой результат, получаемый для одного из классов, на все другие классы ансамбля, изменив лишь соответствующий параметр θ вероятностной меры класса. Поскольку больше всего результатов на сегодня получено для равновероятных подстановок ($\theta_i = 1$, $m_i = (i-1)!$, $i = 1, \ldots, n$), все эти результаты можно перенести с соответствующими переформулировками на любые другие классы комбинаторных объектов ансамбля с параметрами $\theta_i' = \dfrac{(i-1)!}{m_i'}$, $i = 1, \ldots, n$. То же самое и в случае схемы Эванса: $\theta_i' = \dfrac{(i-1)!}{m_i'}\theta$, $i = 1, \ldots, n$.

Подчеркнем, что отмеченное свойство «переноса» имеет место для комбинаторных объектов лишь типа ансамбля.

3. Добавим к предыдущему еще один асимптотический результат для подстановок в общей параметрической схеме.

Пусть $n \to \infty$ и существует единственный ограниченный корень $z_n = z_n(\theta)$ уравнения

$$\sum_{i \in R} z^i \theta_i = n \tag{22}$$

(точка перевала знаменателя в (19)). Тогда из (19) следует, что при фиксированном $r \geqslant 1$

$$\mathbf{E}_\theta (c_i)_r \sim \left(\frac{z_n^i}{i}\theta_i\right)^r,$$

т. е. при ограниченном θ_i распределение числа $c_i(n)$ циклов фиксированной длины $i \in R$ будет асимптотически пуассоновским с параметром $\dfrac{z_n^i}{i}\theta_i$.

В частности, если $R = X_n$ и $\theta_i = b\theta^i$, $i \geqslant 1$, то $z_n \sim \theta^{-1}$ и

$$L_\theta\left(c_i\left(n\right)\right) \sim \Pi\Big(\frac{b}{i}\Big)$$

(предельное распределение не зависит от θ).

Если же $R = \{1, 2, \ldots, m\}$ и $\ln m \asymp \ln n$ (по-прежнему $\theta_i = b\theta^i$), то

$$z_n = \frac{1}{\theta} + \frac{1}{m\theta}\left(\ln\frac{n}{m} + \ln\ln\frac{n}{m} - \ln b + o\left(1\right)\right),$$

и общее число циклов случайной R-подстановки будет асимптотически нормальным со средним и дисперсией

$$\sum_{i\in R}\frac{z_n^i}{i}.$$

4. Двупараметрическая модель на множестве подстановок. Пусть $\theta > 0$ и целое m, $1 \leqslant m < n$, — два заданных параметра. Положим в (1) $\theta_i = \binom{m}{i}\theta$, $i = 1, \ldots, m$, $\theta_i = 0$, $i = m+1, \ldots, n$.

Таким образом, параметр m определяет подмножество длин допустимых циклов в рассматриваемой схеме, а параметр θ определяет вероятностную меру на этом подмножестве.

Уравнение (18) точки перевала в этом случае принимает вид

$$\theta\sum_{i=1}^{m}\binom{m}{i}z^i = n,$$

или

$$(z+1)^m - 1 = \frac{n}{\theta},$$

откуда находим точное решение

$$z_n = \Big(\frac{n}{\theta} + 1\Big)^{1/m} - 1.$$

Распределение числа $c_i(n)$ циклов фиксированной длины определяется поведением величины z_n при $n \to \infty$ в зависимости от поведения параметров m и θ, которые могут быть выбраны как функции от n: $\theta = \theta(n)$ и $m = m(n)$. В частности, если m конечно, то распределение числа $c_i(n)$ циклов фиксированной длины $i \leqslant m$, будет асимптотически нормальным с легко вычислимыми по формулам (19) средними и дисперсиями.

СПИСОК ЛИТЕРАТУРЫ

1. *Ивченко Г. И., Медведев Ю. И.* Неравновероятные меры на множествах разложимых комбинаторных объектов. — Обозрение прикл. промышл. матем., 2003, т. 10, в. 2, с. 348–349.

2. *Ивченко Г. И., Медведев Ю. И.* Случайные комбинаторные объекты. — Докл. РАН, 2004, т. 396, № 2, с. 151–154.

3. *Ивченко Г.И., Медведев Ю.И., Сачков В.Н.* Некоторые проблемы вероятностной комбинаторики. — В сб.: Математика и безопасность информационных технологий. Материалы конференции в МГУ 23.10.2003. — М.: МЦНМО, 2004, с. 45–52.
4. *Колчин В.Ф.* Случайные отображения. — М.: Наука,1984.
5. *Колчин В.Ф.* Случайные графы. — М.: ФИЗМАТЛИТ, 2000.
6. *Сачков В.Н.* Введение в комбинаторные методы дискретной математики. — М.: МЦНМО, 2004.
7. *Ewens W.J.* The sampling theory of selectively neutral alleles. — Theor. Popul. Biol., 1972, v. 3, p. 87–112.
8. *Arratia R., Tavare S.* Independent process approximations for random combinatorial structures. — Adv. Math., 1994, v. 104, p. 90–154.
9. *Arratia R., Barbour A.D., Tavare S.* On Poisson-Dirichlet limits for random decomposable combinatorial structures. — Comb. Probab. Comp., 1999, v. 8, № 3, p. 193–208.
10. *Ивченко Г.И., Медведев Ю.И.* О случайных подстановках. — В сб.: Труды по дискретной математике. Т. 5. — М.: ФИЗМАТЛИТ, 2002, с. 73–92.
11. *Ивченко Г.И., Медведев Ю.И.* Метод В. Л. Гончарова и его развитие в анализе различных моделей случайных подстановок. — Теор. вероятн. примен., 2002, т. 47, в. 3, с. 558–566.
12. *Ивченко Г.И., Медведев Ю.И.* Об одном классе неравновероятных подстановок случайной степени. — В сб.: Труды по дискретной математике. Т. 7. — М.: ФИЗМАТЛИТ, 2003, с. 75–88.