



Math-Net.Ru

All Russian mathematical portal

G. I. Ivchenko, Yu. I. Medvedev, Parametric models for random combinatorial objects of exponential type and their probabilistic-statistical analysis, *Mat. Vopr. Kriptogr.*, 2017, Volume 8, Issue 3, 41–56

DOI: 10.4213/mvk231

Use of the all-Russian mathematical portal Math-Net.Ru implies that you have read and agreed to these terms of use

<http://www.mathnet.ru/eng/agreement>

Download details:

IP: 18.97.14.84

February 9, 2025, 12:18:24



УДК 519.212.2

**Параметрические модели случайных комбинаторных  
объектов экспоненциального типа и вопросы их  
вероятностно-статистического анализа**

**Г. И. Ивченко, Ю. И. Медведев**

*Академия криптографии Российской Федерации, Москва*

*Получено 30.V.2016*

**Аннотация.** Работа посвящена исследованию некоторых проблем вероятностной комбинаторики, когда на множестве рассматриваемых комбинаторных объектов задается некоторая параметрическая вероятностная мера. Проводится сравнительный анализ асимптотических свойств случайных  $n$ -подстановок и случайных разбиений  $n$ -множеств в однопараметрической модели.

**Ключевые слова:** случайные комбинаторные объекты, параметрическая модель, подстановки, разбиения конечных множеств, предельные теоремы, статистические выводы

**Parametric models for random combinatorial objects of exponential type and their probabilistic-statistical analysis**

**G. I. Ivchenko, Yu. I. Medvedev**

*Academy of Cryptography of the Russian Federation, Moscow*

**Abstract.** Some problems of probabilistic combinatorics are investigated when on the set of combinatorial objects under consideration some parametric probabilistic measure is defined. We compare asymptotic properties of random partitions of an  $n$ -set and of  $n$ -permutations in a parametric model is performed.

**Key words:** random combinatorial objects, parametric model, permutations, partitions of finite sets, limit theorems, statistical inferences

Citation: *Mathematical Aspects of Cryptography*, 2017, v. 8, № 3, pp. 41–56 (Russian)

© Академия криптографии Российской Федерации, 2017 г.

## 1. Введение

В множестве случайных комбинаторных объектов по их аналитическим и теоретико-вероятностным свойствам можно выделить три типа: экспоненциальные, биномиальные и отрицательно-биномиальные [7, 8, 11]. Предметом данного исследования будут объекты экспоненциального типа, к которым относятся, в частности, подстановки и разбиения конечных множеств. Параметрические модели случайных подстановок уже детально рассматривались в [10, 12]. Что касается случайных разбиений конечных множеств, то известные результаты здесь [1–5] относятся к равновероятной модели, когда на множестве всех разбиений  $n$ -множества  $X_n = \{1, 2, \dots, n\}$  задается равномерное вероятностное распределение, приписывающее каждому разбиению вероятность  $1/T_n$ , где  $T_n$  (общее число разбиений  $n$ -множества) есть известные в комбинаторике числа Белла [14]. Для теории и приложений представляет интерес рассмотрение и других моделей разбиений конечных множеств, допускающих те или иные отклонения от равновероятности. Некоторые классы таких неравновероятных моделей и будут предметом настоящего исследования. При этом мы будем проводить параллель с уже хорошо изученным объектом — случайными подстановками, что представляется методологически самоценным, поскольку, как отмечено выше, оба эти объекта — подстановки и разбиения — относятся к одному типу комбинаторных объектов.

Общая параметрическая модель для разложимых (т. е. обладающих свойством разложимости на отдельные компоненты) комбинаторных объектов была введена и исследована в работах [7, 8, 11]. Напомним кратко необходимые для последующего изложения основные элементы этой теории.

Пусть  $n$  есть *характеристический параметр* рассматриваемых объектов, их **вес** (для отображений и разбиений конечного множества  $n$  — число элементов множества), и пусть  $\mathbf{K}_n = \{K_n\}$  есть совокупность всех таких объектов веса  $n$ . Для конкретного объекта  $K_n \in \mathbf{K}_n$  обозначим через  $c_i(n)$  число его компонент веса  $i$  ( $i = 1, 2, \dots, n$ ). Набор  $c(n) = (c_1(n), c_2(n), \dots, c_n(n))$ ,  $\sum_{i=1}^n i c_i(n) = n$ , называется структурой объекта  $K_n$ . Наконец, пусть  $N(n, c(n))$  обозначает число объектов  $K_n \in \mathbf{K}_n$  со структурой  $c(n)$ .

На множестве  $\mathbf{K}_n = \{K_n\}$  задается вероятностная мера, зависящая от многомерного параметра, в соответствии с которой объект  $K_n$  со структурой  $c(n) = a = (a_1, a_2, \dots, a_n)$  наблюдается с вероятностью, пропорциональной  $\prod_i \theta_i^{a_i}$ .

Именно,

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

где  $I(\cdot)$  — индикатор и

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

есть нормирующий множитель, представляющий собой производящую функцию чисел  $N(n, a)$  и называемый *статистической суммой* множества  $\mathbf{K}_n$ .

Для объектов экспоненциального типа числа  $N(n, a)$  и статистическая сумма  $H_n(\theta)$  имеют следующие представления:

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

и

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

где  $m_i$  есть число объектов веса  $i$ , состоящих из единственной компоненты (здесь и далее используется обозначение  $[z^n]f(z) = \text{coef}_{z^n} f(z)$ ).

Числа  $m_i$  для многих объектов в комбинаторике хорошо известны. Так, для подстановок  $m_i = (i - 1)!$ , а для разбиений конечных множеств  $m_i = 1$ , поэтому для подстановок и разбиений статистические суммы имеют вид соответственно

$$H_n^{(n)}(\theta) = n! [z^n] \exp\left\{\sum_{i=1}^n \frac{z^i}{i} \theta_i\right\} \tag{1}$$

и

$$H_n^{(p)}(\theta) = n! [z^n] \exp\left\{\sum_{i=1}^n \frac{z^i}{i!} \theta_i\right\}. \tag{2}$$

В рассматриваемой общей модели производящая функция распределения структуры  $c(n) = (c_1(n), c_2(n), \dots, c_n(n))$  случайного объекта имеет вид

$$A_{n,\theta}(t_1, \dots, t_n) = \mathbf{E}_\theta \prod_i t_i^{c_i(n)} = \frac{H_n(t \bullet \theta)}{H_n(\theta)}, \tag{3}$$

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

Отметим некоторые частные случаи этих общих соотношений для разбиений (для подстановок аналогичные соотношения отмечались ранее в цитированных выше работах). Далее будем использовать традиционное в литературе обозначение  $\kappa_n = (\kappa_n(1), \kappa_n(2), \dots, \kappa_n(n))$  ( $\kappa_n(i)$  — число блоков размера  $i$  в случайном разбиении  $n$ -множества) для структуры разбиений, а для цикловой структуры подстановок сохраним обозначение  $c(n) = (c_1(n), c_2(n), \dots, c_n(n))$  (здесь  $c_i(n)$  — число циклов длины  $i$  в случайной  $n$ -подстановке).

1) Имеет место следующая общая формула для смешанных факториальных моментов случайной структуры  $\kappa_n$  (функция  $H_n^{(p)}(\theta)$  определена формулой (2) и  $(a)_r = a(a-1)\dots(a-r+1)$ ,  $r \geq 1$ ,  $(a)_0 = 1$ ):

$$\mathbf{E}_\theta \prod_i (\kappa_n(i))_{r_i} = \frac{n!}{(n-m)!} \frac{H_{n-m}^{(p)}(\theta)}{H_n^{(p)}(\theta)} \prod_i \left(\frac{\theta_i}{i!}\right)^{r_i}, \quad m = \sum_i i r_i \leq n.$$

В частности, формулы для средних значений имеют вид

$$\mathbf{E}_\theta \kappa_n(i) = \frac{n!}{(n-i)!} \frac{\theta_i H_{n-i}^{(p)}(\theta)}{i! H_n^{(p)}(\theta)}.$$

Совместное распределение величин  $\kappa_n(1), \dots, \kappa_n(n)$  может быть представлено как условное распределение:

$$\mathcal{L}(\kappa_n(i), i = 1, \dots, n) = \mathcal{L}\left(Z_1, \dots, Z_n \mid \sum_{i=1}^n i Z_i = n\right),$$

где случайные величины  $Z_1, \dots, Z_n$  независимы в совокупности и имеют пуассоновские распределения с параметрами соответственно  $\lambda_i = \theta_i/i!$ ,  $i = 1, 2, \dots, n$ .

2) При  $\theta_i = 1$ ,  $i = 1, \dots, n$ , получаем равномерную меру на множестве  $\mathbf{K}_n$  всех разбиений множества  $X_n$ , приписывающую каждому разбиению  $K_n \in \mathbf{K}_n$  одну и ту же вероятность  $\mathbf{P}(K_n) = 1/T_n$ , поскольку (см. (2))

$$\begin{aligned} H_n^{(p)}(1) &= n! [z^n] \exp \left\{ \sum_{i=1}^n \frac{z^i}{i!} \right\} = n! [z^n] \exp \left\{ \sum_{i=1}^{\infty} \frac{z^i}{i!} \right\} = \\ &= n! [z^n] \exp \{e^z - 1\} = T_n. \end{aligned}$$

Последнее равенство в этой цепочке есть известное определение чисел Белла  $T_n$  (см., например, [5]).

3) Пусть  $A$  есть некоторое подмножество множества натуральных чисел, и символ  $\mathbf{K}_n(A)$  обозначает подмножество тех разбиений  $K_n \in \mathbf{K}_n$ , размеры блоков которых являются элементами  $A$  (т.е. рассматриваются так называемые  $A$ -разбиения [5]).

Полагая  $\theta_i = 0$  при  $i \notin A$ , получим меру, сосредоточенную на  $\mathbf{K}_n(A)$ .

4) При  $\theta_1 = \theta_2 = \dots = \theta_n = \theta > 0$  получаем однопараметрическую меру, которая приписывает разбиениям  $K_n \in \mathbf{K}_n$  вероятности, пропорциональные  $\theta^{\xi(n)}$ , где  $\xi(n) = \sum_{i=1}^n \kappa_n(i)$  — общее число блоков разбиения  $K_n$ . В этом случае, как следует из (2),

$$\begin{aligned} H_n^{(p)}(\theta) &= n! [z^n] \exp \{ \theta (e^z - 1) \} = e^{-\theta} \sum_{k=1}^{\infty} \frac{k^n \theta^k}{k!} = \\ &= \sum_{k=0}^n \sigma(n, k) \theta^k \equiv B_n(\theta) \end{aligned} \tag{4}$$

есть полином Белла [14], а  $\sigma(n, k)$  — число Стирлинга 2-го рода, дающее число разбиений  $n$ -множества на  $k$  блоков.

Производящая функция случайной величины  $\xi(n)$  равна, следовательно, отношению двух полиномов Белла:

$$\mathbf{E}_\theta t^{\xi(n)} = P_{n,\theta}(t) = A_{n,\theta}(t, \dots, t) = B_n(t\theta) / B_n(\theta), \tag{5}$$

а производящая функция совместного распределения случайных величин  $\kappa_n(i)$ ,  $i = 1, 2, \dots, k$ ,  $k \geq 1$ , имеет вид

$$\begin{aligned} \mathbf{E}_\theta \prod_{i=1}^k t_i^{\kappa_n(i)} &= A_{n,\theta}(t_1, \dots, t_k, 1, \dots, 1) = \\ &= \frac{n!}{B_n(\theta)} [z^n] \exp \left\{ \theta \left( \sum_{i=1}^k (t_i - 1) \frac{z^i}{i!} + e^z - 1 \right) \right\}. \end{aligned}$$

**Замечание.** Для сравнения укажем, что соответствующий результат для случайных  $n$ -подстановок имеет следующий вид [10]: статистическая сумма есть (см. (1))

$$\begin{aligned} H_n^{(n)}(\theta) &= n! [z^n] \exp \{ -\theta \ln(1 - z) \} = n! [z^n] (1 - z)^{-\theta} = \\ &= \theta(\theta + 1) \dots (\theta + n - 1) \equiv [\theta]_n = \sum_{k=1}^n |s(n, k)| \theta^k, \end{aligned}$$

где  $s(n, k)$  — числа Стирлинга первого рода, производящая функция общего числа циклов  $C_s(n) = \sum_{i=1}^n c_i(n)$  случайной подстановки  $s \in S_n$  есть

$$\mathbf{E}_\theta t^{C_s(n)} = \frac{[\theta t]_n}{[\theta]_n},$$

и в этом случае величина  $|s(n, k)|$  есть число  $n$ -подстановок с  $k$  циклами.

5) Вообще, можно рассматривать произвольное конечное разбиение  $\bigcup_{r=1}^N A_r = X_n$  и изучать совместное распределение чисел  $A_r$ -блоков, т.е. тех блоков, размеры которых являются элементами  $A_r : \xi_r(n) = \sum_{i \in A_r} \kappa_n(i)$ , в следующей  $N$ -параметрической модели:  $\theta_i = \vartheta_r, i \in A_r, r = 1, 2, \dots, N$ . Производящая функция их совместного распределения имеет вид

$$\mathbf{E}_\theta \prod_{r=1}^N t_r^{\xi_r(n)} = \frac{h_n(t_1 \vartheta_1, \dots, t_N \vartheta_N)}{h_n(\vartheta_1, \dots, \vartheta_N)}, \quad (6)$$

где

$$h_n(x_1, \dots, x_N) = [z^n] \exp \left\{ \sum_{r=1}^N x_r \sum_{i \in A_r} \frac{z^i}{i!} \right\}.$$

Далее детально рассматривается одномерная параметрическая модель: проведен ее теоретико-вероятностный и статистический анализ, включая оценивание неизвестного параметра и построение соответствующих статистических критериев. Отдельно рассматривается случай, когда объем множества точно не известен, но есть информация о его распределении как случайной величины.

## 2. Однопараметрическая модель

В этом разделе рассматривается случай однопараметрической меры, описанный в п. 4) введения. В этом случае общее число блоков разбиения  $K_n$ , т.е.  $\xi(n) = \sum_{i=1}^n \kappa_n(i)$ , является достаточной статистикой модели, распределение которой, как следует из соотношений (4) и (5), имеет вид

$$\mathbf{P}_\theta \{ \xi(n) = k \} = \frac{\theta^k \sigma(n, k)}{B_n(\theta)}, \quad k = 0, 1, \dots, n.$$

Отсюда, в свою очередь, следует, что уравнение

$$\mathbf{E}_\theta \varphi(\xi(n)) = \frac{1}{B_n(\theta)} \sum_{k=0}^n \varphi(k) \sigma(n, k) \theta^k = 0 \quad \forall \theta > 0$$

имеет единственное решение  $\varphi(k) = 0, k = 0, 1, \dots, n$ , что означает полноту статистики  $\xi(n)$ .

1) *Оценивание.* Рассмотрим сначала задачу построения несмещенных оценок для параметрических функций  $\tau(\theta)$ . Как известно, при наличии полной достаточной статистики несмещенные оценки можно искать лишь в классе функций от этой статистики, т.е. в нашем случае в классе функций вида  $\varphi(\xi(n))$ . Таким образом, общий вид уравнения несмещенности в рассматриваемой модели есть

$$\mathbf{E}_\theta \varphi(\xi(n)) = \frac{1}{B_n(\theta)} \sum_{k=0}^n \varphi(k) \sigma(n, k) \theta^k = \tau(\theta) \quad \forall \theta > 0.$$

Если это уравнение при заданной функции  $\tau(\theta)$  имеет решение (такие функции от параметра называются *оцениваемыми*), то решение единственно и определяет несмещенную оценку с минимальной дисперсией для функции  $\tau(\theta)$ . Такую оценку мы будем называть *оптимальной* и обозначать символом  $\tau^* = \tau^*(\xi(n))$ .

Но при любой функции  $\varphi$  выражение  $\sum_{k=0}^n \varphi(k) \sigma(n, k) \theta^k$  есть многочлен от  $\theta$  степени не выше  $n$ , обращающийся в ноль при  $\theta = 0$ . Отсюда следует, что оцениваемыми в нашей модели являются лишь функции вида  $\tau(\theta) = a(\theta)/B_n(\theta)$ , где  $a(\theta)$  — многочлен степени не выше  $n$ , обращающийся в ноль при  $\theta = 0$ .

Далее, если  $a(\theta) = \sum_{k=1}^n a_k \theta^k$ , то из тождественного равенства

$$\sum_{k=1}^n \varphi(k) \sigma(n, k) \theta^k = \sum_{k=1}^n a_k \theta^k, \quad \theta > 0,$$

следует, что

$$\varphi(k) = \frac{a_k}{\sigma(n, k)}, \quad k = 1, 2, \dots, n.$$

Таким образом, оптимальная оценка для функции  $\tau(\theta) = a(\theta)/B_n(\theta)$  имеет вид

$$\tau^* = \varphi(\xi(n)) = \frac{a_{\xi(n)}}{\sigma(n, \xi(n))}.$$

Подчеркнем, что для функций  $\tau(\theta) \equiv \theta$  и  $\tau(\theta) = \theta^{-1}$  несмещенных оценок не существует.

2) *Распределение числа блоков случайного разбиения.* Производящая функция  $E_\theta t^{\xi(n)} = P_{n,\theta}(t)$  числа блоков случайного разбиения  $K_n$ , указанная в (5), представляет собой отношение двух полиномов Белла (см. (4)). Известно [15], что все корни полинома Белла  $B_n(\theta)$  различные, действительные и неположительные. Поэтому, обозначив его ненулевые корни через  $-\alpha_1, -\alpha_2, \dots, -\alpha_{n-1}$ , можно записать представление

$$B_n(\theta) = \theta(\theta + \alpha_1) \dots (\theta + \alpha_{n-1}).$$

Отсюда следует, что

$$P_{n,\theta}(t) = t \prod_{j=1}^{n-1} \left( \frac{t\theta}{\theta + \alpha_j} + \frac{\alpha_j}{\theta + \alpha_j} \right). \tag{7}$$



Но это означает, что случайная величина  $\xi(n)$  может быть представлена суммой независимых бернуллиевских случайных величин:

$$\xi(n) = \xi_{n1} + \xi_{n2} + \dots + \xi_{n,n-1} + 1, \quad (8)$$

где

$$p_{nj}(\theta) = \mathbf{P} \{ \xi_{nj} = 1 \} = 1 - \mathbf{P} \{ \xi_{nj} = 0 \} = \frac{\theta}{\theta + \alpha_j}, \quad j = 1, 2, \dots, n-1.$$

Отсюда также следуют представления для ее среднего значения и дисперсии:

$$\begin{aligned} \mathbf{E}_\theta \xi(n) &= 1 + \sum_{j=1}^{n-1} p_{nj}(\theta) = 1 + \theta \sum_{j=1}^{n-1} \frac{1}{\theta + \alpha_j}, \\ \mathbf{D}_\theta \xi(n) &= \sum_{j=1}^{n-1} p_{nj}(\theta) (1 - p_{nj}(\theta)) = \theta \sum_{j=1}^{n-1} \frac{\alpha_j}{(\theta + \alpha_j)^2}. \end{aligned} \quad (9)$$

Представление (8) с использованием стандартных методов (см. [13]) позволяет установить для числа блоков  $\xi(n)$  как интегральный, так и локальный варианты центральной предельной теоремы с оценкой скорости сходимости в ней.

**Теорема 1.** Если при  $n \rightarrow \infty$  выполняется условие  $\mathbf{D}_\theta \xi(n) \rightarrow \infty$ , то имеет место соотношение

$$\sup_x \left| \mathbf{P}_\theta \left\{ \frac{\xi(n) - \mathbf{E}_\theta \xi(n)}{\sqrt{\mathbf{D}_\theta \xi(n)}} < x \right\} - \Phi(x) \right| \leq \frac{C}{\sqrt{\mathbf{D}_\theta \xi(n)}}, \quad (10)$$

где  $\Phi(x)$  — стандартная нормальная функция распределения и  $C$  — некоторая положительная константа, а также

$$\max_k \left| \sqrt{\mathbf{D}_\theta \xi(n)} \mathbf{P}_\theta \{ \xi(n) = k \} - \frac{1}{\sqrt{2\pi}} e^{-x^2(k)/2} \right| \leq \frac{C}{\sqrt{\mathbf{D}_\theta \xi(n)}}, \quad (11)$$

где

$$x(k) = \frac{k - \mathbf{E}_\theta \xi(n)}{\sqrt{\mathbf{D}_\theta \xi(n)}}.$$

Дополним эту теорему некоторыми комментариями.

Асимптотическая нормальность с параметрами  $(n/\ln n, n/(\ln n)^2)$  числа блоков в случайном равновероятном разбиении  $n$ -множества, т. е. при гипотезе  $H_0: \theta = 1$ , впервые была доказана в [15] и впоследствии неоднократно передоказывалась другими способами (см., например, [5, с. 308–313]), а скорость сходимости с оценкой вида  $O(\ln n/\sqrt{n})$  в нормальной предельной теореме получена в [16]. При гипотезе  $H_0$  для моментов  $\mathbf{E}_1 \xi(n)$  и  $\mathbf{D}_1 \xi(n)$  и их асимптотик известны следующие результаты [5]:

$$\mathbf{E}_1 \xi(n) = \frac{T_{n+1}}{T_n} - 1 = \frac{n}{\ln n} (1 + o(1)),$$

$$\mathbf{D}_1 \xi(n) = \frac{T_{n+2}}{T_n} - \left(\frac{T_{n+1}}{T_n}\right)^2 - 1 = \frac{n}{(\ln n)^2} (1 + o(1)).$$

С учетом этого из соотношений (9) легко получить, что при «близких» (при  $n \rightarrow \infty$ ) однопараметрических альтернативах вида

$$H_{1n}: \theta = \theta_n = 1 + \delta_n, \quad \delta_n \rightarrow 0,$$

справедливы представления

$$\mathbf{E}_{\theta_n} \xi(n) = 1 + \theta_n \sum_{j=1}^{n-1} \frac{1}{\theta_n + \alpha_j} = \mathbf{E}_1 \xi(n) + \delta_n \mathbf{D}_1 \xi(n) (1 + O(\delta_n)),$$

$$\mathbf{D}_{\theta_n} \xi(n) \sim \mathbf{D}_1 \xi(n) \sim \frac{n}{(\ln n)^2}.$$

**Замечание.** Можно показать, что при произвольном фиксированном значении параметра  $\theta$  имеют место асимптотические представления

$$\mathbf{E}_\theta \xi(n) = \frac{n}{\ln n} \left( 1 + \frac{\ln \ln n}{\ln n} + \frac{\ln \theta}{\ln n} + O\left(\left(\frac{\ln \ln n}{\ln n}\right)^2\right) \right),$$

$$\mathbf{D}_\theta \xi(n) = \frac{n}{(\ln n)^2} \left( 1 + 2 \frac{\ln \ln n}{\ln n} + 2 \frac{\ln \theta}{\ln n} + O\left(\left(\frac{\ln \ln n}{\ln n}\right)^2\right) \right).$$

Таким образом, утверждения теоремы 1 являются новыми: они распространяют классические результаты на неравновероятные модели случайных разбиений и позволяют построить и рассчитать статистический критерий проверки гипотезы  $H_0: \theta = 1$  с учетом таких альтернатив.

Подобные критерии могут представлять интерес в ситуациях, когда при разбиении множества одни типы разбиения являются более предпочтительными, другие — менее. Например, если более предпочтительным является появление большего числа блоков, то следует выбирать параметрическую модель с  $\theta > 1$ , и наоборот, при предпочтении малого числа блоков следует выбирать модель с параметром  $\theta < 1$ .

3) *Статистический критерий*. Пусть, для определенности, «близкие» альтернативы  $H_{1n}$  задаются параметром  $\delta_n < 0$ . Тогда, основываясь на теореме 1, стандартными рассуждениями получаем (рутинные выкладки опускаем), что асимптотически при  $\delta_n \rightarrow 0$  критерий уровня значимости  $\alpha$  для проверки гипотезы  $H_0$  при таких альтернативах  $H_{1n}$  задается критической областью

$$X_\alpha(n) = \left\{ \xi(n) < \mathbf{E}_1 \xi(n) - u_\alpha \sqrt{\mathbf{D}_1 \xi(n)} \right\}, \quad u_\alpha = \Phi^{-1}(1 - \alpha). \quad (12)$$

Для мощности  $W_n(\theta_n) = \mathbf{P}_{\theta_n} \{X_\alpha(n)\}$  этого критерия при альтернативе  $H_{1n}$  имеем:

$$\begin{aligned} W_n(\theta_n) &= \mathbf{P}_{\theta_n} \left\{ \frac{\xi(n) - \mathbf{E}_{\theta_n} \xi(n)}{\sqrt{\mathbf{D}_{\theta_n} \xi(n)}} < \frac{\mathbf{E}_1 \xi(n) - \mathbf{E}_{\theta_n} \xi(n)}{\sqrt{\mathbf{D}_{\theta_n} \xi(n)}} - u_\alpha \sqrt{\frac{\mathbf{D}_1 \xi(n)}{\mathbf{D}_{\theta_n} \xi(n)}} \right\} = \\ &= \mathbf{P}_{\theta_n} \left\{ \frac{\xi(n) - \mathbf{E}_{\theta_n} \xi(n)}{\sqrt{\mathbf{D}_{\theta_n} \xi(n)}} < -\delta_n \sqrt{\mathbf{D}_1 \xi(n)} (1 + o(1)) - u_\alpha (1 + o(1)) \right\} \rightarrow \\ &\rightarrow \begin{cases} 1, & \text{если } -\delta_n \frac{\sqrt{n}}{\ln n} \rightarrow +\infty, \\ \Phi(t - u_\alpha), & \text{если } -\delta_n \frac{\sqrt{n}}{\ln n} \rightarrow t > 0, \\ \alpha, & \text{если } -\delta_n \frac{\sqrt{n}}{\ln n} \rightarrow 0. \end{cases} \end{aligned}$$

Здесь условием  $-\delta_n \frac{\sqrt{n}}{\ln n} \rightarrow +\infty$  определяются так называемые «пороговые» альтернативы (т. е. такие, для которых асимптотическое значение мощности критерия больше  $\alpha$ , но меньше 1): они имеют вид

$$H_{1n}: \theta = \theta_n = 1 - t \frac{\ln n}{\sqrt{n}}, \quad t > 0,$$

и при таких альтернативах предельное значение мощности критерия равно  $\Phi(t - u_\alpha)$ .

Более близкие альтернативы, т. е. такие, для которых выполняется условие  $\delta_n = o((\ln n)/\sqrt{n})$ , критерий (12) асимптотически от гипотезы  $H_0: \theta = 1$  не отличает, а против более далеких альтернатив, т. е. удовлетворяющих условию  $-\delta_n \frac{\sqrt{n}}{\ln n} \rightarrow +\infty$ , он обладает свойством состоятельности.

Для сравнения отметим, что общее число  $C(n)$  циклов случайной  $n$ -подстановки также имеет представление (8), но теперь  $\xi_{ni}$  — бернуллиевская случайная величина, принимающая значения 0 и 1 с вероятностями соответственно  $q_i$  и  $p_i$ :  $p_i = 1 - q_i = \theta / (\theta + i)$ . Первые два момента этой случайной величины имеют асимптотическое (при  $n \rightarrow \infty$ ) представление

$$\mathbf{E}_\theta C(n) \sim \mathbf{D}_\theta C(n) \sim \theta \ln n,$$

и для ее распределения также справедливы соотношения (10) – (11) (с заменой  $\xi(n)$  на символ  $C(n)$ ).

Что же касается задачи проверки статистической гипотезы о равновероятности подстановок в соответствующей параметрической модели, то «пороговыми» альтернативами являются здесь альтернативы вида [9]

$$H_{1n}: \theta = \theta_n = 1 - \frac{t}{\sqrt{\ln n}}, \quad t > 0.$$

### 3. Множество случайного объема

В ряде прикладных задач точный объем исследуемого множества неизвестен, и можно лишь предполагать априори его распределение. По этой причине представляется целесообразным проведение анализа модели в предположении, что объем множества является случайной величиной. Будем обозначать ее символом  $\eta$  и рассмотрим случай, когда она имеет распределение типа степенного ряда:

$$\mathbf{P}_\theta \{ \eta = n \} = \frac{x^n [x^n] f_\theta(x)}{f_\theta(x)}, \quad n = 0, 1, 2, \dots \tag{13}$$

Если функция  $f_\theta(x)$  выбрана в виде

$$f_\theta(x) = \exp \left\{ \theta \sum_{i=1}^{\infty} \frac{x^i}{i!} \right\} = \exp \{ \theta (e^x - 1) \}, \quad x > 0 \tag{14}$$

(согласно (4) это есть экспоненциальная производящая функция полиномов Белла), то  $\kappa_\eta(i)$ ,  $i = 1, 2, \dots$ , являются (см. [11]) независимыми пуассоновскими случайными величинами:

$$\mathcal{L}(\kappa_\eta(i)) = \Pi \left( \theta \frac{x^i}{i!} \right), \quad i \geq 1. \tag{15}$$

В этой модели  $x > 0$  является свободным параметром, выбором которого можно распорядиться в зависимости от цели исследования. Подчеркивая зависимость изучаемых характеристик от этого параметра, в дальнейшем будем снабжать их соответствующей меткой, например для объема множества далее будем использовать символ  $\eta(x)$ .

1) *О распределении объема  $\eta(x)$  множества.* Из (13) для производящей функции  $\eta(x)$  следует представление

$$\mathbf{E}_\theta z^{\eta(x)} = \frac{f_\theta(zx)}{f_\theta(x)} = \exp\{\theta e^x (e^{(z-1)x} - 1)\}, \quad (16)$$

откуда для среднего значения объема множества и его дисперсии имеем

$$\mathbf{E}_\theta \eta(x) = \mu_\theta(x) = x \frac{f'_\theta(x)}{f_\theta(x)} = \theta x e^x, \quad \mathbf{D}_\theta \eta(x) = \sigma_\theta^2(x) = \theta x (1+x) e^x. \quad (17)$$

Пусть теперь  $x \rightarrow \infty$ . Тогда из (16) и (17) стандартными рассуждениями нетрудно вывести, что для характеристической функции нормированной случайной величины  $\eta^*(x) = (\eta(x) - \mu_\theta(x)) / \sigma_\theta(x)$  выполняется предельное соотношение

$$\mathbf{E}_\theta \exp\{\eta^*(x)\} \rightarrow e^{-t^2/2},$$

означающее, что случайная величина  $\eta(x)$  асимптотически нормальна с параметрами  $(\mu_\theta(x), \sigma_\theta^2(x))$ .

2) *О распределении общего числа блоков  $\xi(x) = \sum_{i \geq 1} \kappa_\eta(i)$ .* Эта характеристика, как следует из (15), в силу независимости слагаемых имеет пуассоновское распределение с параметром  $\theta(e^x - 1)$ .

В частности, с вероятностью

$$\mathbf{P}_\theta \{\xi(x) = 0\} = \exp\{-\theta(e^x - 1)\}$$

в такой модели может появиться пустое множество; чтобы вероятность такого неблагоприятного события была мала, надо рассматривать лишь случай  $x \rightarrow \infty$ , и тогда для числа блоков будет иметь место асимптотическая нормальность:

$$\mathcal{L}(\xi(x)) \sim \mathcal{N}(\theta(e^x - 1), \theta(e^x - 1)).$$

Также при  $x \rightarrow \infty$  будут асимптотически нормальны и все величины  $\kappa_\eta(i)$ ,  $i \geq 1$ :

$$\mathcal{L}(\kappa_\eta(i)) \sim \mathcal{N}\left(\theta \frac{x^i}{i!}, \theta \frac{x^i}{i!}\right).$$

3) *Объем максимального блока  $\mu_\theta(x) = \max\{i: \kappa_\eta(i) > 0\}$ .* Для распределения этой характеристики можно записать следующую цепочку очевидных соотношений:

$$\mathbf{P}_\theta \{\mu_\theta(x) \leq s\} = \mathbf{P}_\theta \{\kappa_\eta(i) = 0, i > s\} = \prod_{i > s} \mathbf{P}_\theta \{\kappa_\eta(i) = 0\} = e^{-\theta E(s, x)}, \quad (18)$$

где  $E(s, x) = \sum_{i > s} x^i / i!$ .

Из (18) следует, что дальнейшее исследование распределения величины  $\mu_\theta(x)$  требует анализа поведения функции  $E(s, x)$ .

Будем предполагать, что  $x \rightarrow \infty$ . Тогда асимптотически распределение  $\mu_\theta(x)$  будет сосредоточено в зоне таких значений  $s = s(x)$ , для которых выполняется соотношение  $E(s, x) \rightarrow \text{const} > 0$ . Условие выполнения этого соотношения приведено в следующей лемме.

**Лемма.** Пусть при  $x \rightarrow \infty$  величина  $s = s(x)$  меняется таким образом, что

$$\frac{x^s}{s!} \rightarrow \alpha = \text{const} > 0. \tag{19}$$

Тогда  $\alpha = 1/\sqrt{2\pi}$ ,  $s = ex(1 - \frac{\ln x}{2ex}(1 + o(1)))$  и

$$E(s, x) \rightarrow \frac{1}{(e-1)\sqrt{2\pi}}. \tag{20}$$

*Доказательство.* Из разложения

$$E(s, x) = \sum_{i>s} \frac{x^i}{i!} = \frac{x^s}{s!} \left( \frac{x}{s+1} + \frac{x^2}{(s+1)(s+2)} + \dots \right)$$

следует, что если выполняется (19), то  $E(s, x) \rightarrow \alpha\gamma(1-\gamma)^{-1}$ , где  $\gamma = \lim \frac{x}{s}$ .

В свою очередь, используя формулу Стирлинга, соотношение (19) можно переписать в виде

$$\left(\frac{xe}{s}\right)^s \frac{1}{\sqrt{2\pi s}} \rightarrow \alpha, \quad \text{или} \quad s \ln \frac{xe}{s} - \ln \sqrt{2\pi s} \rightarrow \ln \alpha.$$

Поэтому, во-первых,  $\gamma = \lim \frac{x}{s} = e^{-1}$ , а во-вторых,

$$\frac{xe}{s} = 1 + \frac{\ln s}{2s} \left(1 + o\left(\frac{1}{\ln s}\right)\right).$$

Отсюда уже следуют как вид константы в (19) и асимптотика для  $s = s(x)$ , указанные в лемме, так и окончательное соотношение (20).  $\square$

Эта лемма позволяет легко установить вид предельного распределения объема максимального блока в рассматриваемой модели.

**Теорема 2.** Пусть  $x \rightarrow \infty$  и величина  $s = s(x)$  меняется так, как указано в (19). Тогда для любого фиксированного  $k = 0, \pm 1, \pm 2, \dots$

$$\mathbf{P}_\theta \{ \mu_\theta(x) \leq s + k \} \rightarrow \exp \left\{ -\theta \frac{e^{-k}}{(e-1)\sqrt{2\pi}} \right\}.$$

*Доказательство.* Действительно, согласно (18) и лемме (в ее обозначениях)

$$\begin{aligned} \mathbf{P}_\theta \{ \mu_\theta(x) \leq s+k \} &= \exp \{ -\theta E(s+k, x) \} = \\ &= \exp \left\{ -\theta \frac{x^{s+k}}{(s+k)!} \left( \frac{x}{s+k+1} + \frac{x^2}{(s+k+1)(s+k+2)} + \dots \right) \right\} \rightarrow \\ &\rightarrow \exp \left\{ -\theta \alpha \gamma^{k+1} (1-\gamma)^{-1} \right\}, \end{aligned}$$

что дает указанный в формулировке результат.  $\square$

Таким образом, при больших значениях  $x$  распределение  $\mu_\theta(x)$  сосредоточено в окрестности точки  $s = ex \left( 1 - \frac{\ln x}{2ex} (1 + o(1)) \right)$  и является распределением дискретного типа.

4) *Сравнения.* Интересно сравнить этот результат с аналогичным результатом для классической модели (т. е. для  $\theta = 1$ ) с объемом множества (см. (17))  $n = xe^x = \mathbf{E}_1 \eta(x)$ . Решение этого уравнения имеет, как хорошо известно, вид

$$x = \ln n - \ln \ln n + O\left(\frac{\ln \ln n}{\ln n}\right), \quad n \rightarrow \infty, \quad (21)$$

так что распределение величины  $\mu_1(x)$  сосредоточено в окрестности точки  $e \ln n$ .

В то же время для среднего значения величины максимального блока  $\mu_n$  в случайном разбиении  $n$ -множества при  $n \rightarrow \infty$  имеет место асимптотика (см. [2, с. 160–161])

$$\mathbf{E} \mu_n = er - \ln \sqrt{r} + O(1),$$

где  $re^r = n$  (т. е. это среднее также имеет порядок  $e \ln n$ ), а асимптотическое поведение распределения  $\mu_n$  описывается следующим образом.

Введем случайную величину

$$\tilde{\mu}_n = \mu_n - \left[ er - \ln \sqrt{2\pi er} - \ln(e-1) \right],$$

тогда равномерно для всех таких  $x$ , что  $|x| < \ln \sqrt{r}$ ,

$$\mathbf{P} \{ \tilde{\mu}_n < x \} = \exp \left\{ -e^{-[x]+\delta} \right\} \rightarrow 0,$$

где  $\delta = \{ er - \ln \sqrt{2\pi er} - \ln(e-1) \}$  — дробная часть соответствующего числа.

Таким образом, при больших значениях  $n$  распределение  $\mu_n$  сосредоточено в окрестности точки  $\left[ er - \ln \sqrt{2\pi er} - \ln(e-1) \right]$  и близко к дважды экспоненциальному типу.

Что касается асимптотического поведения общего числа блоков  $\xi(x)$ , то при  $x$  вида (21) эта величина является асимптотически нормальной с параметрами  $(\theta n / \ln n, \theta n / (\ln n)^2)$ , что при  $\theta = 1$  совпадает с известным аналогичным результатом для  $n$ -множеств (см., например, [4]).

Также (при  $\theta = 1$  и условии (21)) в обеих моделях одинаковы и асимптотические распределения величин  $\kappa_\eta(i)$ ,  $i \geq 1$ : они оказываются независимыми и нормальными с параметрами  $(x^i / i!, x^i / i!)$  [1].

Аналогичная модель с рандомизацией объема множества рассматривалась и для случайных подстановок [6]. В этом случае соответствующий результат (аналог теоремы 2) для длины максимального цикла случайной подстановки имеет принципиально иной характер: при соответствующей нормировке она имеет предельную функцию распределения вида

$$e^{-\theta E(u)}, \quad \text{где} \quad E(u) = \int_1^\infty e^{-uy} \frac{dy}{y}.$$

В заключение выражаем благодарность А. М. Зубкову за внимательное прочтение рукописи и высказанные замечания, способствовавшие улучшению изложения ее содержания.

## Список литературы

- [1] Сачков В. Н., “Случайные разбиения множеств”, *Теория вероятн. и ее примен.*, **19**:1 (1974), 178–194.
- [2] Сачков В. Н., *Вероятностные методы в комбинаторном анализе*, М.: Наука, 1978, 288 с.
- [3] Сачков В. Н., “Случайные разбиения множеств”, *Математические вопросы кибернетики*, **8**, 1999, 33–54.
- [4] Сачков В. Н., “Разбиения с поглощениями и противоречивые разбиения множеств”, *Труды по дискретной математике*, **4**, 2001, 201–222.
- [5] Сачков В. Н., *Введение в комбинаторные методы дискретной математики. 2-е изд.*, М.: МЦНМО, 1978, 424 с.
- [6] Ивченко Г. И., Медведев Ю. И., “Об одном классе неравновероятных подстановок случайной степени”, *Труды по дискретной математике*, **7**, 2003, 75–88.
- [7] Ивченко Г. И., Медведев Ю. И., Сачков В. Н., “Некоторые проблемы вероятностной комбинаторики”, В сб.: *Математика и безопасность инф. технологий. Матер. конф.*, М.: МЦНМО, 2004, 45–52.
- [8] Ивченко Г. И., Медведев Ю. И., “Случайные комбинаторные объекты”, *Доклады АН*, **396**:2 (2004), 151–154.
- [9] Ивченко Г. И., Медведев Ю. И., “Статистика параметрической модели случайных подстановок”, *Труды по дискретной математике*, **8**, 2004, 116–127.
- [10] Ивченко Г. И., Медведев Ю. И., “Случайные подстановки: общая параметрическая модель”, *Дискретная математика*, **18**:4 (2006), 105–112.



- [11] Ивченко Г. И., Медведев Ю. И., “Случайные комбинаторные объекты в общей параметрической модели”, Труды по дискретной математике, **10**, 2007, 73–86.
- [12] Ивченко Г. И., Соболева М. В., “Некоторые неравновероятные модели случайных подстановок”, *Дискретная математика*, **23**:3 (2011), 23–31.
- [13] Вагутин В. А., Михайлов В. Г., “Предельные теоремы для числа пустых ячеек в равновероятной схеме размещения частиц комплектами”, *Теория вероятн. и ее примен.*, **27**:4 (1982), 684–692.
- [14] Bell E. T., “Exponential polynomials”, *Ann. Math.*, **35**:2 (1934), 258–277.
- [15] Harper L., “Stirling behavior is asymptotically normal”, *Ann. Math. Stat.*, **38**:2 (1967), 410–414.
- [16] Hwang H. K., “On convergence rates in the central limit theorems for combinatorial structures”, *Eur. J. Combin.*, **19**:3 (1998), 329–343.