ОБОЗРЕНИЕ
ПРИКЛАДНОЙ И ПРОМЫШЛЕННОЙ
Том 19 МАТЕМАТИКИ Выпуск 4
2012

Г. И. Ивченко, **М. В. Солдаткина** (Москва, МИЭМ НИУ ВШЭ). **Статистические задачи для случайных подстановок с цензурированными данными.**

Рассматривается следующая d-параметрическая модель случайных n-подстановок, т. е. взаимно однозначных отображений множества $X_n = \{1, 2, \ldots, n\}$ в себя. Пусть задано некоторое разбиение множества X_n:

$$X_n = \bigcup_{j=1}^{d} A_j, A_i \bigcap A_j = \varnothing, \qquad i \neq j, \tag{1}$$

и $\mathbf{c}(n) = (c_1, c_2, \ldots, c_n)$ есть цикловая структура n-подстановки s (c_i-число ее циклов длины i, $i = 1, \ldots, n$, $\sum_i ic_i = n$). Циклы подстановки s, длины которых являются элементами подмножества A_j называются A_j-циклами, их число обозначается $C_{A_j}(n) = \sum_{i \in A_j} c_i$, $j = 1, \ldots, d$. Если подмножество A_j имеет вид

$$A_j = \{k : k = ld + j,\ l \geqslant 0\}, \tag{2}$$

для некоторых целых $d \geqslant 2$ и $1 \leqslant j \leqslant d$, то говорят о *конгруэнтных* циклах [1, с. 187].

Далее, на множестве всех n-подстановок $S_n = \{s\}$ задается параметрическая вероятностная мера вида

$$\mathbf{P}_\theta(s) = I\left(\sum_{i=1}^{n} ic_i = n\right) \prod_{j=1}^{d} \theta_j^{C_{A_j}(n)} \Big/ H_n(\theta), \tag{3}$$

где $I(\cdot)$ — индикатор, $\theta = (\theta_1, \ldots, \theta_d)$, $\theta_j \geqslant 0$, $j = 1, \ldots, d$, — произвольные параметры, не равные нулю одновременно, и $H_n(\theta)$-необходимый нормирующий множитель, имеющий вид

$$H_n(\theta) = n![z^n] \exp\left\{\sum_{j=1}^{d} \theta_j \sum_{i \in A_j} \frac{z^i}{i}\right\} \tag{4}$$

(здесь и далее $[z^n]f(z) = coef_{z^n} f(z)$).

В [2] доказана асимптотическая (при $n \to \infty$) нормальность вектора $\mathbf{C}(n) = (C_{A_1}(n), \ldots, C_{A_d}(n))$ чисел конгруэнтных циклов случайной подстановки в такой модели и решены соответствующие задачи проверки статистических гипотез о параметре $\theta = (\theta_1, \ldots, \theta_d)$. Вопросы оценивания неизвестного параметра θ по наблюдению над вектором $\mathbf{C}(n)$ рассмотрены в работе [3].

В этих исследованиях предполагается известной вся цикловая последовательность $\mathbf{c}(n) = (c_1, c_2, \ldots, c_n)$, т. е. имеется *полная информация* о наблюдаемой подстановке, и соответсвующие выводы имеют асимптотический (при $n \to \infty$) характер. Но в этом случае не всегда является реалистичным предположение о том, что мы можем наблюдать всю цикловую последовательность $\mathbf{c}(n)$. Может быть и так, что наблюдению доступно лишь какое-то ограниченное число k ее первых членов $c_1, c_2, \ldots, c_k$, — в этом случае говорят о *цензурированных* (*неполных*) *данных.*

В настоящем докладе предполагается, что в наблюдаемой подстановке $s \in S_n$ для каждого $j = 1, \ldots, d$ доступно подсчету лишь число A_j-циклов с длинами, не превосходящими заданного уровня K_j. В таком случае, пусть

$$\xi_j(n) = \sum_{i \in A_j} c_i I(i \leqslant K_j), \qquad j = 1, \ldots, d, \tag{5}$$

есть наши исходные данные (количества наблюдаемых A_j-циклов). Рассматриваются различные вопросы статистического вывода для модели (1)–(4) именно по таким неполным данным (5). При этом порядок подстановки $n \to \infty$, а параметры цензурирования K_j, $j = 1, \ldots, d$, фиксированы.

Для начальных членов цикловой структуры $\mathbf{c}(n)$ подстановки s установлен следующий результат.

Теорема. *Для случайной подстановки s в модели* (1) – (4) *при $n \to \infty$ числа циклов ограниченной длины асимптотически независимы, и при этом число A_j-циклов длины i имеет в пределе распределение Пуассона* $\Pi(\theta_j/i)$.

Следствие. *Наблюдаемые статистики* (5) *асимптотически независимы, и при этом*

$$(\xi_j(n)) \to \Pi(\theta_j \lambda_j),$$
$$\text{где } \lambda_j = \sum_{i \in A_j} \frac{1}{i} I(i \leqslant K_j) = \sum_{l \leqslant (K_j - j)/d} \frac{1}{ld + j}. \tag{6}$$

Из этих результатов следует, во-первых, что статистические выводы о каждом из параметров θ_j можно делать независимо по наблюдению лишь соответсвующей статистики $\xi_j(n)$, и, во-вторых, исходная проблема в асимптотике сводится к соответствующим статистическим задачам для пуассоновской модели с неизвестным параметром, решение которых достаточно хорошо известно.Таким образом, полученный результат позволяет решать задачи точечного и доверительного оценивания параметров θ_j d-параметрической модели случайных n-подстановок с неполными данными и задачи проверки статистических гипотез.

СПИСОК ЛИТЕРАТУРЫ

1. *Сачков В. Н.* Введение в комбинаторные методы дискретной математики. М.: МЦ НМО, 2004.
2. *Соболева М. В.* Асимптотическая нормальность чисел конгруэнтных циклов в случайных подстановках. — Дискретн. матем., 2012, т. 24, в. 1, с. 123–131.
3. *Солдаткина М. В.* Оценивание параметров в одной модели случайных подстановок. — Обозрение прикл. и промышл. матем., 2012, т. 19, в. 2, с. 222–223.
4. *Ивченко Г. И.*, *Медведев Ю. И.* Статистика параметрической модели случайных подстановок. — Труды по дискретной математике. Т. 8. М.: Физматлит, 2004, с. 116–127.
5. *Ивченко Г. И.*, *Медведев Ю. И.* Статистические выводы для случайных подстановок по неполным данным. Труды по дискретной математике. Т. 9. М.: Физматлит, 2006, с. 66–76.
6. *Ивченко Г. И.*, *Медведев Ю. И.* Случайные подстановки: общая параметрическая модель. Дискретн. матем., 2006, т. 18, в. 4, с. 105–112.
7. *Ивченко Г. И.*, *Медведев Ю. И.* Введение в математическую статистику. М: ЛКИ/URSS, 2010.