Труды Карельского научного центра РАН
№ 5. 2012. С. 106–109

УДК 519.2

ОЦЕНИВАНИЕ ПАРАМЕТРОВ В ОДНОЙ МОДЕЛИ СЛУЧАЙНЫХ ПОДСТАНОВОК

М. В. Солдаткина

Московский институт электроники и математики НИУ ВШЭ

Рассматривается d-мерная параметрическая модель случайных n-подстановок, для которой при $n \to \infty$ устанавлена совместная асимптотическая нормальность чисел конгруэнтных циклов в случайной подстановке. Изучаются вопросы оценивания параметров $\theta = (\theta_1, \ldots, \theta_d)$ модели по наблюдению над вектором цикловой структуры.

Ключевые слова: случайные подстановки, конгруэнтные циклы, параметрическая модель, асимптотическое оценивание параметров.

M. V. Soldatkina. PARAMETER ESTIMATION FOR A RANDOM PERMUTATIONS MODEL

In the context of a d-dimensional parametric model of random n-permutations, as $n \to \infty$, we established the joint asymptotic normality of the numbers of congruent cycles in a random permutation. We study the issues of estimating the parameters $\theta = (\theta_1, \ldots, \theta_d)$ of the model for observation of the cycle structure vector.

Key words: random permutations, congruent cycles, parametric model, asymptotic parameter estimation.

Введение

Тематика случайных подстановок, берущая начало с работы В. Л. Гончарова [1], весьма многопланова и продолжает постоянно расширяться. Для последнего времени характерен переход от рассмотрения равновероятных подстановок к построению более общих моделей случайных подстановок с различного рода отклонениями от равновероятности. Весьма общая параметрическая модель такого рода была введена в работах Г. И. Ивченко и Ю. И. Медведева [2, 3], согласно которой произвольная подстановка s из симметрической группы S_n подстановок n-й степени наблюдается с вероятностью, пропорциональной $\prod_i \theta_j^{c_i}$, где c_i – число циклов длины i в подстановке s ($i = 1, \ldots, n$, $\sum\limits_i ic_i = n$) и $\theta = (\theta_1, \ldots \theta_n)$, $\theta_. \geqslant 0$, – параметр соответствующей меры $\mathbf{P}_\theta$ на S_n .

Для практических применений представляется естественным более детально исследовать различные конкретизации этой общей модели при тех или иных ограничениях на число степеней свободы меры $\mathbf{P}_\theta$. Таким задачам были посвящены работы [5, 6], в которых предложены новые параметрические меры, подходящие для изучения цикловой структуры A-подстановок и их обобщений.

Напомним, что для заданного подмножества $A \subset X_n = \{1, 2, \ldots, n\}$ A-циклами подстановки $s \in S_n$ называются те ее циклы, длины которых являются элементами A. Если задано некоторое разбиение

$$X_n = \bigcup_{j=1}^{d} A_j, A_j \bigcap A_k = \varnothing, j \neq k, d \geqslant 2, \quad (1)$$

и $C_{A_j}(n) = \sum_{i \in A_j} c_i$ есть число A_j-циклов подстановки s, $j = 1, \ldots, d$, то для исследования обобщенной цикловой структуры (будем говорить об $A = (A_1, ..., A_d)$-структуре) подстановки s, т. е. вектора

$$\mathbf{C}_A(n) = (C_{A_1}(n), C_{A_2}(n), \ldots C_{A_d}(n)), \quad (2)$$

на S_n вводится вероятностная мера вида

$$\mathbf{P}_{A\theta}(s) = \mathbf{I}\left(\sum_{i=1}^{n} ic_i = n\right) \prod_{j=1}^{d} \theta_j^{C_{A_j}(n)} \Big/ H_{An}(\theta), \theta = (\theta_1, \ldots \theta_d), \quad (3)$$

где $\mathbf{I}(\cdot)$ – индикатор и $H_{An}(\theta)$ – необходимый нормирующий множитель, имеющий вид:

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

(здесь и далее $[z^n]f(z) = coef_{z^n} f(z)$). Как показано в работе [5], производящая функция вектора $\mathbf{C}_A(n)$ в модели (3)–(4) имеет вид:

$$\mathbf{E}_\theta \prod_{j=1}^{d} t_i^{C_{A_j}(n)}$$

$$= H_n(t \times \theta)/H_n(\theta), t \times \theta$$

$$= (t_1\theta_1, \ldots, t_d\theta_d).$$

Далее в работе [6] эта конструкция была применена для исследования чисел конгруэнтных циклов в подстановке, т. е. когда подмножества A_j имеют вид:

$$A_j = \{k : k = ld + j, l \geqslant 0\} \quad (5)$$

для некоторых целых $d \geqslant 2$ и $1 \leqslant j \leqslant d$, и был установлен следующий результат.

Теорема 1. *Если $n \to \infty$, а параметры $\theta_1, \ldots, \theta_d$ фиксированы, то компоненты вектора (2) асимптотически независимы и асимптотически нормальны с параметрами соответственно $\left(\frac{\theta_j}{d}\ln n, \frac{\theta_j}{d}\ln n\right)$, $j = 1, \ldots, d$, при этом параметры нормальных распределений являются асимптотическими значениями соответствующих средних и дисперсий компонент А-структуры, и эта асимптотика равномерна по $\theta = (\theta_1, \ldots \theta_d)$ в любой конечной области изменения параметра.*

Этот результат позволяет решать и естественные статистические задачи оценивания параметров, и проверки гипотез в рамках рассматриваемой модели – тематика, мало исследованная для случайных подстановок. Частично эти вопросы затронуты в работе [6], где предложен новый статистический критерий для проверки гипотезы о равновероятности подстановок $H_0 : \theta_1 = \ldots = \theta_d = 1$ с вычислением его мощности для близких альтернатив.

Настоящая работа является продолжением этих исследований: она посвящена вопросам оценивания параметров $\theta = (\theta_1, \ldots, \theta_d)$ модели (3)–(4) по наблюдению над вектором A-структуры (2).

Асимптотическое оценивание

Возможности точного анализа в задачах оценивания для рассматриваемой модели довольно ограничены, как это констатируется в работе [4], где рассматривался случай однопараметрической модели. Именно в [4] обсуждаются вопросы оценивания и проверки гипотез о параметре θ модели Эванса, когда произвольная подстановка $s \in S_n$ наблюдается с вероятностью, пропорциональной $\theta^{c_s(n)}$, где $c_s(n)$ – общее число циклов подстановки s и $\theta > 0$ – неизвестный параметр. Это частный случай нашей модели при $d = 1$ и $A_1 = X_n$. Для него в [4] показано, что несмещённые оценки существуют лишь для параметрических функций вида $\tau(\theta) = a(\theta)/\theta_{(n)}$, где $\theta_{(n)} = \theta(\theta+1)\ldots(\theta+n-1)$ и $a(\theta)$ – многочлен степени не выше n, причём $a(0) = 0$. В частности, для самого параметра θ несмещённой оценки не существует. В [4] построены также оптимальные (несмещённые с минимальной дисперсией) оценки для таких параметрических функций, при этом базой для этого является известное точное распределение достаточной статистики $c_s(n)$. Также констатируется, что более широкие возможности в обсуждаемой проблематике предоставляет асимптотический подход, когда порядок подстановок $n \to \infty$, и для этого случая развивается соответствующая асимптотическая теория, в рамках которой конструируются асимптотически несмещенные и асимптотически эффективные оценки не только для параметра θ, но и для широкого класса функций от него, а также рассчитываются соответствующие асимптотические доверительные интервалы.

В свете сказанного, при анализе нашей модели (3)–(4) мы будем использовать асимптотический подход ($n \to \infty$), поскольку точное распределение достаточной статистики $\mathbf{C}_A(n)$ в удобной для использования форме получить весьма проблематично, в то время как её асимптотическое распределение, указанное в теореме, достаточно просто устроено.

Из приведённой теоремы следует, что если ввести нормированные статистики

$$\tilde{C}_{A_j}(n) = \frac{dC_{A_j}(n)}{\ln n}, j = 1, \ldots, d, \qquad (6)$$

то они будут асимптотически независимы и нормальны:

$$\mathcal{L}(\tilde{C}_{A_j}(n)) \sim \mathcal{N}\left(\theta_j, \frac{d\theta_j}{\ln n}\right), j = 1, \ldots, d. \qquad (7)$$

Более того, на основании теорем сходимости для функций от случайных величин [7, с. 335] можно заключить, что для любых дифференцируемых функций $\tau_j(\theta)$ имеет место аналогичное утверждение для статистики $\tau_j(\tilde{C}_{A_j}(n))$:

$$\mathcal{L}(\tau_j(\tilde{C}_{A_j}(n))) \qquad (8)$$
$$\sim \mathcal{N}\left(\tau_j(\theta_j), (\tau_j^{'}(\theta_j))^2 \frac{d\theta_j}{\ln n}\right), j = 1, \ldots, d.$$

На основании соотношений (7) и (8) можно сформулировать следующее утверждение.

Утверждение 1. *Если $n \to \infty$, то статистика $\tilde{C}_{A_j}(n)$(статистика $\tau_j(\tilde{C}_{A_j}(n))$) является асимптотически несмещенной и асимптотически эффективной оценкой для параметра θ_j(для дифференцируемой параметрической функции $\tau_j(\theta_j)$), и параметры $\theta_1, \ldots, \theta_d$ оцениваются независимо друг от друга.*

Соотношение (7) позволяет также рассчитать и асимтотический доверительный интервал для параметра θ_j. Для этого прежде всего заметим, что (на основании теорем сходимости [7]) соотношение (7) останется справедливым, если в его правой части заменить параметр θ_j, входящий в выражение дисперсии, оценкой $\tilde{C}_{A_j}(n)$. Таким образом, при $n \to \infty$ справедливо также соотношение:

$$\mathcal{L}\left(\frac{\tilde{C}_{A_j}(n) - \theta_j}{\sqrt{d\tilde{C}_{A_j}(n)}} \sqrt{\ln n}\right) \sim \mathcal{N}(0,1). \qquad (9)$$

Пусть теперь задан доверительный уровень γ, $0 < \gamma < 1$. Определим число z_γ уравнением $\Phi(z_\gamma) = \frac{1+\gamma}{2}$, где $\Phi(z)$ – стандартная нормальная функция распределения. Тогда из (9) следует, что

$$\mathbf{P}\left\{-z_\gamma < \frac{\tilde{C}_{A_j}(n) - \theta_j}{\sqrt{d\tilde{C}_{A_j}(n)}} \sqrt{\ln n} < z_\gamma\right\}$$
$$= \mathbf{P}\left\{\tilde{C}_{A_j}(n) - z_\gamma \sqrt{\frac{d\tilde{C}_{A_j}(n)}{\ln n}} < \theta_j \right.$$
$$\left. < \tilde{C}_{A_j}(n) + z_\gamma \sqrt{\frac{d\tilde{C}_{A_j}(n)}{\ln n}}\right\}$$
$$\sim \Phi(z_\gamma) - \Phi(-z_\gamma) = 2\Phi(z_\gamma) - 1 = \gamma.$$

Это означает, что справедливо

Утверждение 2. *Асимптотический $\gamma-$доверительный интервал для параметра θ_j имеет вид:*

$$\left(\tilde{C}_{A_j}(n) \mp z_\gamma \sqrt{d\tilde{C}_{A_j}/\ln n}\right).$$

Аналогично, на основании соотношения (8) строится асимптотический доверительный интервал для функции $\tau_j(\theta_j)$. Однако, чтобы получить в этом случае аналог соотношения (9) (с подстановкой в выражение диспресии статистики $\tilde{C}_{A_j}(n)$ вместо неизвестного параметра θ_j), необходимо дополнительно потребовать, чтобы производная $\tau_j^{'}(\theta_j)$ была непрерывна [7]. При выполнении этого условия искомый γ-доверительный интервал асимтотически имеет вид:

$$\left(\tau_j(\tilde{C}_{A_j}(n)) \mp z_\gamma \acute{\tau}_j(\tilde{C}_{A_j}(n)) \sqrt{d\tilde{C}_{A_j}(n)/\ln n}\right).$$

В частности, для функции $\tau_j(\theta_j) = \sqrt{\theta_j}$ соответствующий интервал особенно прост:

$$\left(\sqrt{\tilde{C}_{A_j}(n)} \mp \frac{z_\gamma}{2} \sqrt{\frac{d}{\ln n}}\right).$$

Многовыборочный случай

Пусть в модели (3)–(4) наблюдается $N \geqslant 2$ независимых подстановок при одном и том же (но неизвестном) значении параметра $\theta = (\theta_1, \ldots \theta_d)$, и $\mathbf{C}_A^{(i)}(n) = (C_{A_1}^{(i)}(n), C_{A_2}^{(i)}(n), \ldots C_{A_d}^{(i)}(n))$ есть реализация

A-структуры (2) для i-й подстановки, $i = 1, \ldots, N$.

Введём статистику:

$$\mathbf{T}(N,n) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{C}_A^{(i)}(n)$$

$$= (T_1(N,n), \ldots, T_d(N.n)).$$

В условиях теоремы компоненты этого вектора будут асимптотически независимы и асимптотически нормальны $\mathcal{N}\left(\frac{\theta_j}{d}\ln n, \frac{\theta_j}{dN}\ln n\right)$, $j = 1, \ldots, d$. Нормированные же статистики $\tilde{T}_j(N,n) = dT_j(N,n)/\ln n$ будут асимптотически нормальны $\mathcal{N}\left(\theta_j, \frac{d\theta_j}{N \ln n}\right)$.

Следовательно, $\tilde{T}_j(N,n)$ – асимптотически несмещенная оценка для θ_j с асимптотической дисперсией, в N раз меньшей, чем у оценки $\tilde{C}_{A_j}(n)$ в одновыборочном случае. Таким образом, при таком объединении информации точность оценивания возрастает.

Также более узкими оказываются и соответствующие доверительные интервалы (точность локализации для неизвестных параметров возрастает): асимптотический γ – доверительный интервал для θ_j, основанный на статистике $\tilde{T}_j(N,n)$, имеет вид:

$$(\tilde{T}_j(N,n) \mp z_\gamma \sqrt{d\tilde{T}_j(N,n)/(N \ln n)}).$$

Литература

1. *Гончаров В. Л.* Из области комбинаторики. Изв. АН СССР. Сер. матем. 1944. Т. 8, № 1. С. 3–48.

2. *Ивченко Г. И., Медведев Ю. И.* Случайные комбинаторные объекты // Доклады РАН. 2004. Т. 396, № 2. С. 151–154.

3. *Ивченко Г. И., Медведев Ю. И.* Случайные подстановки: общая параметрическая модель // Дискрет. матем. 2006. Т. 18, № 4. С. 105–112.

4. *Ивченко Г. И., Медведев Ю. И.* Статистика параметрической модели случайных подстановок // Тр. по дискр. матем. 2004. Т. 8. С. 116−127.

5. *Ивченко Г. И., Соболева М. В.* Некоторые неравновероятные модели случайных подстановок // Дискрет. матем. 2011. Т. 23, № 3. С. 23−31.

6. *Соболева М. В.* Асимптотическая нормальность чисел конгруэнтных циклов в случайных подстановках // Дискрет. матем. 2012. Т. 24, № 1. С. 123−131.

7. *Рао С. Р.* Линейные статистические методы и их применения. М.: Наука, 1968. С. 548.

СВЕДЕНИЯ ОБ АВТОРЕ:

Солдаткина Мария Васильевна
аспирантка
Московский институт электроники и математики
НИУ ВШЭ
Б. Трехсвятительский пер., д. 3, Москва, Россия, 109028
эл. почта: manyasha.soboleva@ya.ru
тел.: 8(495) 916 88 29

Soldatkina, Mariya
Moscow State Institute of Electronics and Mathematics
National research university 'Higher school of economics'
3 Bol. Trekhsvjatitel'skij per.,109028, Moscow, Russia
e-mail: manyasha.soboleva@ya.ru
tel.: 8(495) 916 88 29