Math-Net.Ru
All Russian mathematical portal

G. I. Ivchenko, Yu. I. Medvedev, Statistics of a parametric model of random substitutions,
Tr. Diskr. Mat., 2004, Volume 8, 116–127

https://www.mathnet.ru/eng/tdm124

Use of the all-Russian mathematical portal Math-Net.Ru implies that you have read and agreed to these terms of use
https://www.mathnet.ru/eng/agreement

Download details:
IP: 18.97.14.91
January 24, 2026, 05:50:40

Труды по дискретной математике
Том 8, с. 116–127
М.: ФИЗМАТЛИТ, 2004

СТАТИСТИКА ПАРАМЕТРИЧЕСКОЙ МОДЕЛИ СЛУЧАЙНЫХ ПОДСТАНОВОК

Г. И. ИВЧЕНКО, Ю. И. МЕДВЕДЕВ

Для параметрической модели Эванса случайных подстановок решаются задачи несмещенного с минимальной дисперсией оценивания параметрических функций, построения асимптотических (при порядке подстановки $n \to \infty$) доверительных интервалов для параметра и функций от него, а также критериев проверки гипотезы равновероятности и расчета их мощности при «близких» альтернативах.

§ 1. ВВЕДЕНИЕ

Тематика случайных подстановок, берущая начало с класической работы В. Л. Гончарова [1], весьма многопланова и продолжает постоянно расширяться, о чем свидетельствуют и последние публикации на эту тему (см. [2, 6–9] и библиографию в них, а также [3, 4]). Для последних лет характерен переход от рассмотрения равновероятных подстановок к построению более общих моделей случайных подстановок, допускающих те или иные отклонения от равновероятности. Достаточно общей и интересной, как в теоретическом, так и в прикладном аспектах, такой моделью является параметрическая модель Эванса случайных подстановок [10], когда произвольная подстановка A из симметрической группы S_n наблюдается с вероятностью

$$\mathbf{P}_\theta\{A\} = \frac{\theta^{c_A(n)}}{\theta_{(n)}}, \tag{1}$$

где $c_A(n)$ — число циклов подстановки A, $\theta > 0$ — некоторый параметр и

$$\theta_{(n)} = \theta(\theta+1)\dots(\theta+n-1).$$

Если $\theta > 1$, то более вероятными будут подстановки с большим числом циклов (или с короткими циклами), если же $\theta < 1$, то, наоборот, более

© Г. И. Ивченко, Ю. И. Медведев, 2004

вероятными будут подстановки с длинными циклами; при $\theta = 1$ мы имеем случай равновероятных подстановок, так что равновероятная модель является частным случаем модели (1).

Вероятностному анализу модели Эванса (1) посвящен ряд недавних публикаций (библиографию см. в [2]), но статистические аспекты этой модели (т. е. когда параметр θ неизвестен и требуется делать те или иные статистические выводы о нем) до сих пор в литературе не обсуждались, и настоящая работа восполняет этот пробел. В ней рассматриваются задачи построения оптимальных (несмещенных с минимальной дисперсией) точечных оценок параметрических функций, асимптотических (при $n \to \infty$) доверительных интервалов для параметра θ и функций от него, а также проверки гипотезы о равновероятности подстановок в рамках модели (1). Основное внимание уделяется одновыборочному случаю, т. е. когда исходные статистические данные представляют собой цикловую последовательность

$$\bar{c}(n) = (c_1(n), c_2(n), \dots, c_n(n))$$

одной подстановки, наблюдаемой в соответствии с мерой (1). Для этого случая дается полное решение перечисленных выше стандартных статистических задач. Некоторые из этих результатов распространяются также и на многовыборочный случай, когда наблюдается N независимых случайных подстановок.

§ 2. НЕСМЕЩЕННЫЕ ОЦЕНКИ С МИНИМАЛЬНОЙ ДИСПЕРСИЕЙ (Н.О.М.Д)

В этом параграфе рассматривается задача построения точечных оценок для параметрических функций $\tau(\theta)$. При этом мы будем ограничиваться классом несмещенных оценок, т. е. такими статистиками $T(\bar{c}(n))$ (в одновыборочном случае), которые удовлетворяют уравнению несмещенности

$$\mathbf{E}_\theta T(\bar{c}(n)) = \tau(\theta), \qquad \forall \theta > 0. \tag{2}$$

Прежде всего в этом случае возникает задача описания класса оцениваемых функций $\tau(\theta)$, т. е. тех параметрических функций, для которых уравнение (2) имеет решение. Далее для произвольной оцениваемой функции $\tau(\theta)$ мы будем искать оценку с равномерно (по θ) минимальной дисперсией, которую будем обозначать символом

$$\tau^* = \tau^*(\bar{c}(n)).$$

Таким образом, оптимальная н.о.м.д. оценка τ^* для заданной оцениваемой функции $\tau(\theta)$ должна удовлетворять следующим условиям:

$$\mathbf{E}_\theta \tau^* = \tau(\theta) \quad \text{и} \quad \mathbf{D}_\theta \tau^* \leqslant \mathbf{D}_\theta T$$

для всех $\theta > 0$ и любого решения $T = T(\bar{c}(n))$ уравнения (2).

Стандартный метод решения этих задач состоит в отыскании полной достаточной статистики для рассматриваемой модели и использовании теоремы Рао–Блекуэлла–Колмогорова, согласно которой всякая функция от полной достаточной статистики является оптимальной оценкой для своего математического ожидания (см., например, [5, § 2.3]).

Итак, пусть $\bar{c}(n) = (c_1(n), c_2(n), \dots, c_n(n))$ — цикловая последовательность наблюдаемой в модели (1) случайной подстановки, где $c_i(n)$ — число циклов длины i, $i = 1, 2, \dots, n$, и $c(n) = \sum\limits_{i=1}^{n} c_i(n)$ — общее число циклов подстановки. Основой для всех дальнейших рассмотрений является следующее утверждение.

Теорема 1. *Наблюдаемое число циклов $c(n)$ является полной достаточной статистикой для модели* (1), *и ее распределение имеет вид*

$$\mathbf{P}_\theta\{c(n) = k\} = \frac{\theta^k |s(n,k)|}{\theta_{(n)}}, \qquad k = 1, 2, \dots, n, \tag{3}$$

где $s(n,k)$ — числа Стирлинга 1-го рода, определяемые разложением

$$x(x-1)\dots(x-n+1) = \sum_{k=1}^{n} s(n,k)\, x^k.$$

Доказательство. В нашем случае правдоподобие данных (распределение наблюдаемой последовательности $\bar{c}(n)$ при заданном значении параметра θ) имеет вид (см. [10])

$$p_\theta(a_1, \dots, a_n) = \mathbf{P}_\theta\{\bar{c}(n) = \bar{a}\} = \frac{n!}{\theta_{(n)}}\, \theta^{\sum_i a_i} \prod_j \frac{1}{j^{a_j} a_j!}\, I\Big(\sum_i i a_i = n\Big),$$

где $\bar{a} = (a_1, \dots, a_n)$ и $I(A)$ — индикатор события A, т. е. $p_\theta(a_1, \dots, a_n)$ может быть записано в виде

$$p_\theta(a_1, \dots, a_n) = g_\theta\Big(\sum_i a_i\Big)\, h(a_1, \dots, a_n),$$

где функция $h(a_1, \dots, a_n)$ от параметра θ не зависит. Но это по критерию факторизации означает, что статистика $c(n) = \sum\limits_i c_i(n)$ является достаточной. Что касается распределения (3) этой статистики, то оно получено в работе [2]. Для доказательства же ее полноты, как известно, требуется установить, что если функция φ удовлетворяет условию

$$\mathbf{E}_\theta \varphi(c(n)) = 0, \qquad \forall \theta > 0, \tag{4}$$

то $\varphi(k) = 0$ на множестве всех возможных значений статистики $c(n)$, т. е. при $k = 1, 2, \ldots, n$. Но из (3) следует, что условие (4) в данном случае эквивалентно условию

$$\sum_{k=1}^{n} \varphi(k)|s(n,k)|\,\theta^k = 0, \qquad \forall\theta > 0.$$

Из тождественного равенства нулю многочлена следует равенство нулю всех его коэффициентов, а так как $s(n,k) \neq 0$ при $k = 1, 2, \ldots, n$, то, следовательно, $\varphi(k) = 0$ при $k = 1, \ldots, n$. Теорема доказана.

Как известно, при наличии полной достаточной статистики все статистические выводы о неизвестном параметре модели основываются на использовании этой статистики. В частности, решение уравнения несмещенности (2) достаточно искать лишь в классе функций от достаточной статистики, т. е. в нашем случае уравнение несмещенности принимает вид

$$\mathbf{E}_\theta \varphi(c(n)) = \sum_{k=1}^{n} \varphi(k)|s(n,k)|\,\frac{\theta^k}{\theta_{(n)}} = \tau(\theta), \qquad \forall\theta > 0. \tag{5}$$

Но при любой функции φ выражение

$$\sum_{k=1}^{n} \varphi(k)|s(n,k)|\,\theta^k$$

есть многочлен от θ степени, не превосходящей n, обращающийся в нуль при $\theta = 0$. Отсюда следует, что оцениваемыми в рассматриваемой модели являются лишь функции вида $\tau(\theta) = a(\theta)/\theta_{(n)}$, где $a(\theta)$ — многочлен степени не выше n и $a(0) = 0$.

Далее, оптимальная оценка τ^* для заданной оцениваемой параметрической функции $\tau(\theta)$ единственна и является решением уравнения (5), поэтому, если $\tau(\theta) = a(\theta)/\theta_{(n)}$, где $a(\theta) = \sum\limits_{k=1}^{n} a_k\theta^k$, то из тождественного равенства

$$\sum_{k=1}^{n} \varphi(k)|s(n,k)|\,\theta^k = \sum_{k=1}^{n} a_k\theta^k, \qquad \theta > 0,$$

следует, что $\varphi(k) = a_k/|s(n,k)|$, $k = 1, \ldots, n$. Таким образом, оптимальная н.о.м.д. оценка для такой функции $\tau(\theta)$ имеет вид

$$\tau^* = \varphi(c(n)) = \frac{a_{c(n)}}{|s(n, c(n))|}. \tag{6}$$

Сформулируем полученный результат в виде следующего итогового утверждения.

Теорема 2. *В модели* (1) *оцениваемыми являются лишь функции вида* $\tau(\theta) = a(\theta)/\theta_{(n)}$, *где* $a(\theta) = \sum\limits_{k=1}^{n} a_k\theta^k$ *— произвольный многочлен; для такой произвольной функции оптимальной оценкой является статистика* (6).

Выделим некоторые частные случаи этого общего результата.

1) Пусть $a(\theta) = \theta_{(r)} = \theta(\theta+1)\ldots(\theta+r-1)$ при некотором $r \leqslant n$. Тогда

$$\tau(\theta) = \tau_r(\theta) = \frac{\theta_{(r)}}{\theta_{(n)}} = \frac{1}{(\theta+r)\ldots(\theta+n-1)} \tag{7}$$

(при $r = n$ пустое произведение заменяется на 1).

Из определения чисел $s(n, k)$ следует, что

$$\theta_{(r)} = \sum_{k=1}^{r} |s(r, k)|\theta^k,$$

поэтому из (6) следует, что оптимальная оценка для функции вида (7) есть

$$\tau_r^* = \frac{|s(r, c(n))|}{|s(n, c(n))|}. \tag{8}$$

Эта оценка отлична от нуля лишь при $c(n) \leqslant r$.

В частности, при $r = n-1$ статистика

$$\tau_{n-1}^* = \frac{|s(n-1, c(n))|}{|s(n, c(n))|}$$

является оптимальной оценкой для функции $\tau_{n-1}(\theta) = \dfrac{1}{\theta+n-1}$.

2) Если $\tau(\theta) = \theta$, то $a(\theta) = \theta \cdot \theta_{(n)}$ — многочлен степени $n+1$, т. е. для самого параметра θ несмещенной оценки не существует. Также невозможно несмещенно оценить и функцию $\tau(\theta) = \theta^{-1}$, поскольку в этом случае многочлен $a(\theta) = \tau(\theta) \cdot \theta_{(n)} = (\theta+1)\ldots(\theta+n-1)$ не удовлетворяет требованию $a(0) = 0$.

3) Если многочлен $a(\theta)$ задан разложением $a(\theta) = \sum\limits_{r=1}^{n} b_r\theta_{(r)}$, то

$$\tau(\theta) = \frac{a(\theta)}{\theta_{(n)}} = \sum_{r=1}^{n} b_r\tau_r(\theta)$$

(см. (7)) и оптимальная оценка (6) может быть записана в данном случае в виде линейной комбинации оценок (8):

$$\tau^* = \sum_{r=1}^{n} b_r\tau_r^*.$$

§ 3. ДОВЕРИТЕЛЬНОЕ ОЦЕНИВАНИЕ: АСИМПТОТИЧЕСКАЯ ТЕОРИЯ

Как следует из предыдущего, возможности точного анализа в задачах оценивания для модели (1) оказались довольно ограниченными: оптимальные оценки в рассматриваемой модели удается построить лишь для довольно узкого класса параметрических функций, при этом для самого параметра θ модели (1) несмещенная оценка вообще отсутствует. Более широкие возможности в обсуждаемой проблематике предоставляет асимптотический подход, когда порядок подстановок $n \to \infty$. В этом случае удается весьма просто сконструировать асимптотически несмещенные и асимптотически эффективные оценки не только для параметра θ, но и для широкого класса функций от него, а также рассчитать соответствующие асимптотические доверительные интервалы.

Основой для асимптотического анализа является следующее утверждение об асимптотической нормальности числа циклов $c(n)$ (см. [2]): при $\theta > 0$

$$\mathcal{L}_\theta(c(n)) \sim \mathcal{N}(\theta \ln n,\, \theta \ln n), \qquad n \to \infty. \tag{9}$$

Введя нормированное число циклов $\widetilde{c}(n) = c(n)/\ln n$, соотношение (9) можно переписать в виде

$$\mathcal{L}_\theta(\widetilde{c}(n)) \sim \mathcal{N}\left(\theta, \frac{\theta}{\ln n}\right), \qquad n \to \infty. \tag{10}$$

Более того, на основании теорем сходимости для функций от случайных величин (см. [12, с. 335]) можно заключить, что для любой дифференцируемой функции $\tau(\theta)$ имеет место аналогичное асимптотическое соотношение для статистики $\tau(\widetilde{c}(n))$: при $n \to \infty$

$$\mathcal{L}_\theta(\tau(\widetilde{c}(n))) \sim \mathcal{N}\left(\tau(\theta),\, (\tau'(\theta))^2 \frac{\theta}{\ln n}\right). \tag{11}$$

Отметим также, что параметры нормального распределения в (9) являются асимптотическими значениями среднего и дисперсии статистики $c(n)$ соответственно, поэтому соотношение (10) означает, что статистика $\widetilde{c}(n) = c(n)/\ln n$ является асимптотически несмещенной и асимптотически эффективной оценкой для параметра θ. Аналогичное заключение на основании соотношения (11) можно сделать об оценивании дифференцируемых параметрических функций $\tau(\theta)$: оценка в этом случае имеет вид $\tau(\widetilde{c}(n))$.

Соотношение (10) позволяет также рассчитать и асимптотический доверительный интервал для параметра θ. Переписав его в виде

$$\mathcal{L}_\theta\left(\frac{\widetilde{c}(n)-\theta}{\sqrt{\theta}}\sqrt{\ln n}\right)\sim\mathcal{N}(0,1)$$

и заменив здесь в знаменателе параметр θ его оценкой $\widetilde{c}(n)$ (это можно сделать на основании теорем сходимости), получим необходимое соотношение

$$\mathcal{L}_\theta\left(\frac{\widetilde{c}(n)-\theta}{\sqrt{\widetilde{c}(n)}}\sqrt{\ln n}\right)\sim\mathcal{N}(0,1). \tag{12}$$

На основании (12) доверительный интервал для θ строится стандартными рассуждениями. Пусть 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 (12) следует, что

$$\mathbf{P}_\theta\left\{-z_\gamma<\frac{\widetilde{c}(n)-\theta}{\sqrt{\widetilde{c}(n)}}\sqrt{\ln n}<z_\gamma\right\}=$$

$$=\mathbf{P}_\theta\left\{\widetilde{c}(n)-z_\gamma\sqrt{\frac{\widetilde{c}(n)}{\ln n}}<\theta<\widetilde{c}(n)+z_\gamma\sqrt{\frac{\widetilde{c}(n)}{\ln n}}\right\}\sim$$

$$\sim\Phi(z_\gamma)-\Phi(-z_\gamma)=2\Phi(z_\gamma)-1=\gamma.$$

Это и означает, что

$$\left(\widetilde{c}(n)\mp z_\gamma\sqrt{\frac{\widetilde{c}(n)}{\ln n}}\right)$$

есть асимптотический γ-доверительный интервал для неизвестного параметра θ модели (1).

Аналогично, на основании соотношения (11) строится асимптотический доверительный интервал для дифференцируемой параметрической функции $\tau(\theta)$. Если дополнительно предположить, что производная $\tau'(\theta)$ непрерывна, то при доверительном уровне γ искомый интервал имеет вид

$$\left(\tau(\widetilde{c}(n))\mp z_\gamma\tau'(\widetilde{c}(n))\sqrt{\frac{\widetilde{c}(n)}{\ln n}}\right).$$

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

$$\left(\sqrt{\widetilde{c}(n)}\mp\frac{z_\gamma}{2\sqrt{\ln n}}\right).$$

§ 4. ПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ ДЛЯ ГИПОТЕЗЫ О РАВНОВЕРОЯТНОСТИ ПОДСТАНОВОК

Этот параграф посвящен проблеме проверки статистических гипотез о параметре θ модели (1). Мы будем акцентировать внимание на проверке наиболее важной для приложений гипотезы о равновероятности подстановок H_0: $\theta = 1$. Альтернативными гипотезами в данном случае могут быть утверждения H_1^-: $\theta < 1$, H_1^+: $\theta > 1$ и $H_1 = H_1^- \cup H_1^+$: $\theta \neq 1$.

Прежде чем переходить к построению соответствующих критериев, заметим, что в нашей модели отношение правдоподобия

$$L_\theta(a_1, \dots, a_n) = \frac{p_\theta(a_1, \dots, a_n)}{p_1(a_1, \dots, a_n)} = \frac{n!}{\theta_{(n)}}\, \theta^{\sum_i a_i}\, I\Big(\sum_i i a_i = n\Big) \tag{13}$$

при любом $\theta \neq 1$ является монотонной функцией от значения $\sum_i a_i$ достаточной статистики $c(n) = \sum_i c_i(n)$. Следовательно (см. [5, § 4.4]), в задачах (H_0, H_1^-) и (H_0, H_1^+) проверки гипотезы H_0 при односторонних альтернативах H_1^- и H_1^+ равномерно наиболее мощные (р.н.м.) критерии существуют и совпадают с соответствующими критериями Неймана–Пирсона для проверки гипотезы H_0 при простой альтернативе. Более того, критические области соответствующих критериев выражаются через достаточную статистику модели, а из (13) следует, что неравенство $L_\theta(a_1, \dots, a_n) > c$ (задающее критическое множество критерия Неймана–Пирсона) при $\theta < 1$ эквивалентно неравенству $\sum_i a_i < z$, а при $\theta > 1$ — неравенству $\sum_i a_i > z$. Это означает, что в задаче (H_0, H_1^-) проверки гипотезы H_0 против левосторонней альтернативы H_1^- р. н. м. критерий уровня значимости α задается критической областью вида

$$\varkappa_\alpha^-(n) = \{c(n) < z_\alpha^-(n)\}, \tag{14}$$

где критическая граница $z_\alpha^-(n)$ определяется условием

$$\mathbf{P}_1\{\varkappa_\alpha^-(n)\} = \alpha. \tag{15}$$

Аналогичный критерий в задаче (H_0, H_1^+) с правосторонней альтернативой задается критической областью вида

$$\varkappa_\alpha^+(n) = \{c(n) > z_\alpha^+(n)\}, \tag{16}$$

определяемой условием (15), в котором $\varkappa_\alpha^-(n)$ заменено на $\varkappa_\alpha^+(n)$. Наконец, при двусторонней альтернативе H_1 критерий задается объединением двух односторонних областей, т. е. имеет вид (см. [5, § 4.4])

$$\varkappa_\alpha(n) = \varkappa_{\alpha_1}^-(n) \cup \varkappa_{\alpha_2}^+(n), \qquad \alpha_1 + \alpha_2 = \alpha. \tag{17}$$

Для расчета критических границ в (14) и (16) используется асимптотический результат (9), что позволяет в итоге сформулировать асимптотические варианты соответствующих критериев.

Рассмотрим более детально задачу (H_0, H_1^-). Если положить в (14)

$$z_\alpha^-(n) = \ln n - u_\alpha \sqrt{\ln n}, \qquad u_\alpha = \Phi^{-1}(1-\alpha), \tag{18}$$

то при $n \to \infty$

$$\mathbf{P}_1\{\varkappa_\alpha^-(n)\} = \mathbf{P}_1\{c(n) < \ln n - u_\alpha\sqrt{\ln n}\} =$$

$$= \mathbf{P}_1\left\{\frac{c(n) - \ln n}{\sqrt{\ln n}} < -u_\alpha\right\} \to \Phi(-u_\alpha) = 1 - \Phi(u_\alpha) = \alpha,$$

т. е. условие (15) будет выполняться асимптотически. Таким образом, критерий уровня значимости α для проверки гипотезы H_0: $\theta = 1$ при альтернативе H_1^-: $\theta < 1$ имеет вид (14), где при больших n граница $z_\alpha^-(n)$ задается в виде (18): если наблюдается событие $\varkappa_\alpha^-(n)$, то гипотеза H_0 отвергается в пользу альтернативы H_1^-; вероятность ошибочно отвергнуть при этом истинную гипотезу H_0 при больших n приблизительно равна α.

Исследуем мощность $W_n^-(\theta) = \mathbf{P}_\theta\{\varkappa_\alpha^-(n)\}$ этого критерия. При $n \to \infty$

$$W_n^-(\theta) = \mathbf{P}_\theta\left\{\frac{c(n) - \theta\ln n}{\sqrt{\theta\ln n}} < \frac{1-\theta}{\sqrt{\theta}}\sqrt{\ln n} - \frac{u_\alpha}{\sqrt{\theta}}\right\} \to$$

$$\to \begin{cases} 1, & \text{если} \quad (1-\theta)\sqrt{\ln n} \to \infty, \\ \Phi(t - u_\alpha), & \text{если} \quad (1-\theta)\sqrt{\ln n} \to t > 0, \\ \alpha, & \text{если} \quad (1-\theta)\sqrt{\ln n} \to 0. \end{cases}$$

Таким образом, «пороговыми» альтернативами для данного критерия являются альтернативы вида H_{1n}^-: $\theta = \theta_n = 1 - \dfrac{t}{\sqrt{\ln n}}$, $t > 0$; при таких «близких» альтернативах

$$\lim_{n\to\infty} W_n^-(\theta_n) = \Phi(t - u_\alpha). \tag{19}$$

Более близкие альтернативы, т. е. когда $(1-\theta_n)\sqrt{\ln n} \to 0$, данный критерий асимптотически от H_0 не отличает, а против более далеких альтернатив, т. е. когда $(1-\theta_n)\sqrt{\ln n} \to \infty$, он является состоятельным.

Аналогично, в задаче $(H_0,\ H_1^+)$ критерий (16) асимптотически задается границей

$$z_\alpha^+(n) = \ln n + u_\alpha\sqrt{\ln n}, \qquad u_\alpha = \Phi^{-1}(1-\alpha), \tag{20}$$

и его мощность $W_n^+(\theta) = \mathbf{P}_\theta\{\varkappa_\alpha^+(n)\}$ при «близких» альтернативах вида H_{1n}^+: $\theta = \theta_n = 1 + \dfrac{t}{\sqrt{\ln n}}$, $t > 0$, удовлетворяет предельному соотношению

$$\lim_{n\to\infty} W_n^+(\theta_n) = \Phi(t - u_\alpha). \tag{21}$$

Наконец, в задаче с двусторонней альтернативой $(H_0,\ H_1 = H_1^- \cup H_1^+)$ критерий уровня значимости α асимптотически имеет вид (17) с $\alpha_1 = \alpha_2 = \alpha/2$, т. е. (см. (18), (20))

$$\varkappa_\alpha(n) = \left\{\frac{|c(n) - \ln n|}{\sqrt{\ln n}} > u_{\alpha/2}\right\}, \tag{22}$$

и его мощность при «близких» альтернативах вида H_{1n}: $\theta = \theta_n = 1 + \dfrac{t}{\sqrt{\ln n}}$, $t \neq 0$, удовлетворяет предельному соотношению

$$W_n(\theta_n) = \mathbf{P}_{\theta_n}\{\varkappa_\alpha(n)\} \underset{n\to\infty}{\longrightarrow} \Phi(t - u_{\alpha/2}) + \Phi(-t - u_{\alpha/2}). \tag{23}$$

З а м е ч а н и е 1. Известно (см. [11, гл. 4]), что для моделей с монотонным отношением правдоподобия (а наша модель (1) является таковой) р.н.м. критерии для простой гипотезы H_0: $\theta = \theta_0$ при односторонних альтернативах H_1^-: $\theta < \theta_0$ и H_1^+: $\theta > \theta_0$ одновременно являются р. н. м. критериями в задачах $(H_0$: $\theta \geqslant \theta_0,\ H_1^-)$ и $(H_0$: $\theta \leqslant \theta_0,\ H_1^+)$ соответственно. Таким образом, построенный критерий $\varkappa_\alpha^-(n)$ можно использовать и в задаче со сложной односторонней нулевой гипотезой $(H_0$: $\theta \geqslant 1$, H_1^-: $\theta < 1)$; аналогично, критерий $\varkappa_\alpha^+(n)$ можно использовать в задаче $(H_0$: $\theta \leqslant 1$, H_1^+: $\theta > 1)$.

З а м е ч а н и е 2. Выбор в задаче с двусторонней альтернативой H_1: $\theta \neq 1$ «симметричного» критерия (22) обеспечивает важное свойство несмещенности критерия: $W_n(\theta_n) > \alpha$. Нетрудно убедиться, что среди всех критериев вида (17) лишь для критерия (22) предельная мощность в (23) удовлетворяет этому свойству.

§ 5. МНОГОВЫБОРОЧНЫЙ СЛУЧАЙ

Предположим, что наблюдается $N \geqslant 2$ независимых подстановок с распределением (1) при одном и том же (но неизвестном) значении параметра θ. Пусть $c^{(j)}(n)$ обозначает общее число циклов j-й подстановки,

$j = 1, \ldots, N$. Поскольку по предположению параметр θ не менялся в процессе наблюдений, достаточной статистикой для него по всей совокупности данных будет сумма $\sum\limits_{j=1}^{N} c^{(j)}(n)$.

Введем статистику

$$T(N,, n) = \frac{1}{N} \sum_{j=1}^{N} c^{(j)}(n),$$

тогда в силу (9) для нее будет иметь место асимптотическое соотношение

$$\mathcal{L}_\theta(T(N,n)) \sim \mathcal{N}\left(\theta \ln n,\ \theta \frac{\ln n}{N}\right), \qquad n \to \infty,$$

а для нормированной статистики $\widetilde{T}(N,n) = T(N,n)/\ln n$ — аналог соотношения (10):

$$\mathcal{L}_\theta(\widetilde{T}(N,n)) \sim \mathcal{N}\left(\theta,\ \frac{\theta}{N \ln n}\right), \qquad n \to \infty, \tag{24}$$

Из (24) следует, что $\widetilde{T}(N,n)$ — асимптотически несмещенная оценка для θ с асимптотической дисперсией, в N раз меньшей, чем у оценки $\widetilde{c}(n)$ в одновыборочном случае. Таким образом, при таком объединении информации точность оценивания возрастает.

На основании соотношения (24), аналогично § 3, можно получить, что асимптотический γ-доверительный интервал для θ, основанный на статистике $\widetilde{T}(N,n)$, имеет вид

$$\left(\widetilde{T}(N,n) \mp z_\gamma \sqrt{\frac{\widetilde{T}(N,n)}{N \ln n}}\right).$$

Это более узкий интервал, обеспечивающий более точную локализацию неизвестного параметра, нежели интервал, основанный на одновыборочной статистике $\widetilde{c}(n)$.

Использование объединенной статистики $T(N,n)$ в качестве тестовой статистики в обсуждавшихся в § 4 задачах проверки гипотез приводит также к более мощным критериям. Проиллюстрируем это на примере решения задачи (H_0: $\theta = 1$, H_1^-: $\theta < 1$). В этом случае критерий уровня значимости α асимптотически (при $n \to \infty$) имеет вид (сравнить с (14), (18))

$$\varkappa_\alpha^-(N,n) = \left\{T(N,n) < \ln n - u_\alpha \sqrt{\frac{\ln n}{N}}\right\},$$

а его мощность при близких альтернативах

$$H_{1n}^-\colon \theta = \theta_n = 1 - \frac{t}{\sqrt{\ln n}}, \qquad t > 0,$$

удовлетворяет при $n \to \infty$ соотношению (сравнить с (19))

$$W_{Nn}^{-}(\theta_n) = \mathbf{P}_{\theta_n}\{\varkappa_\alpha^{-}(N,n)\} \to \Phi(t\sqrt{N} - u_\alpha).$$

Таким образом, наличие дополнительной информации, суммируемой статистикой $T(N,n)$, позволяет получать более точные статистические выводы.

СПИСОК ЛИТЕРАТУРЫ

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.* О случайных подстановках. — В сб.: Труды по дискретной математике. Т. 5. — М.: ФИЗМАТЛИТ, 2002, с. 73–92.
3. *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.
4. *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.
5. *Ивченко Г. И., Медведев Ю. И.* Математическая статистика. — М.: Высшая школа, 1992.
6. *Сачков В. Н.* Введение в комбинаторные методы дискретной математики. — М.: Наука, 1982.
7. *Сачков В. Н.* Вероятностные методы в комбинаторном анализе. — М.: Наука, 1978.
8. *Колчин В. Ф.* Случайные отображения. — М.: Наука, 1984.
9. *Колчин В. Ф.* Случайные графы. — М.: ФИЗМАТЛИТ, 2000.
10. *Ewens W. J.* The sampling theory of selectively neutral alleles. — Theor. Popul. Biol., 1972, v. 3, p. 87–112.
11. *Леман Э.* Проверка статистических гипотез. — М.: Наука, 1964.
12. *Рао С. Р.* Линейные статистические методы и их применения. — М.: Наука, 1968.