ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

УДК 519.2

О МОЩНОСТИ КРИТЕРИЕВ В СЛУЧАЕ ВЫБОРОК СЛУЧАЙНОГО ОБЪЕМА[1]

Бенинг В.Е.
МГУ им. М.В. Ломоносова, г. Москва
Институт проблем информатики Российской академии наук, г. Москва

Поступила в редакцию 01.11.2018, после переработки 03.12.2018.

В работе получены асимптотические разложения для обобщенных мощностей критериев, основанных на выборках случайного объема. Эти асимптотические разложения тесно связаны с асимптотическим дефектом критериев, который позволяет сравнивать асимптотически эквивалентные критерии.

Ключевые слова: мощность критерия, асимптотическое разложение, асимптотический дефект, выборка случайного объема, трехточечное симметричное распределение.

Вестник ТвГУ. Серия: Прикладная математика. 2018. № 4. С. 5–22.
https://doi.org/10.26456/vtpmk514

1. Введение

Рассмотрим задачу проверки простой гипотезы в случае однопараметрического семейства вероятностных распределений. Пусть имеются независимые наблюдения $\mathbf{X}_n = (X_1, \ldots, X_n)$, каждое из которых принимает значения в произвольном измеримом пространстве $(\mathcal{X}, \mathcal{A})$ и имеет неизвестную с точностью до параметра θ плотность $p(x, \theta)$ относительно некоторой σ–конечной меры на $\mathcal{A}$. Предположим, что неизвестный параметр θ принадлежит открытому множеству $\Theta \subset \mathbb{R}^1$, содержащему ноль. Обозначим через $\mathsf{P}_{n,\theta}$, $\mathsf{E}_{n,\theta}$ соответственно распределение и математическое ожидание $\mathbf{X}_n$, а через $\mathsf{P}_\theta, \mathsf{E}_\theta$ соответственно распределение и математическое ожидание X_1.

Пусть необходимо проверить простую гипотезу

$$\mathsf{H}_0 : \theta = 0 \tag{1.1}$$

против сложной альтернативы $\theta \neq 0$. В общем случае наилучшего (равномерно наиболее мощного) критерия не существует, и поэтому рассматривется асимптотический подход, при котором $n \to \infty$. Рассмотрим сначала простую альтернативу вида (θ_1 известно)

$$\mathsf{H}_1 : \theta = \theta_1 \neq 0. \tag{1.2}$$

[1]Работа выполнена при финансовой поддержке РФФИ (проект № 18-07-00252).

Заметим, что точка $\theta = 0$ в гипотезе H_0 может быть заменена на любую фиксированную точку $\theta_0 \in \Theta$. Этот случай сводится к предыдущему с помощью рассмотрения семейства плотностей вида $p(x, \theta_0 + \xi)$, где ξ принадлежит некоторой окрестности нуля. Согласно фундаментальной лемме Неймана – Пирсона (см., например, [5], теорема 3.2.1) наилучший (наиболее мощный) критерий основан на логарифме отношения правдоподобия

$$\Lambda_n(\theta) = \sum_{i=1}^{n} \big(l(X_i, \theta) - l(X_i, 0)\big), \tag{1.3}$$

где $l(x, \theta) = \log p(x, \theta)$, и отвергает гипотезу H_0 в случае, если

$$\Lambda_n(\theta_1) > c_\alpha(n),$$

причем критическое значение $c_\alpha(n)$ выбирается из условия

$$\mathsf{P}_{n,0}(\Lambda_n(\theta_1) > c_\alpha(n)) = \alpha, \tag{1.4}$$

где $\alpha \in (0, 1)$ – фиксированный уровень значимости, и для простоты предполагается непрерывность распределения $\Lambda_n(\theta_1)$ при гипотезе H_0, то есть считаем, что

$$\mathsf{P}_{n,0}\big(\Lambda_n(\theta_1) = c_\alpha(n)\big) = 0.$$

Далее предположим также, что существуют все необходимые моменты случайных величин $l(X_1, \theta)$ и все необходимые производные по θ функций $l(x, \theta)$.

Обозначим через

$$\mu(\theta) = \mathsf{E}_\theta \,\big(l(X_1, \theta_1) - l(X_1, 0)\big), \tag{1.5}$$

$$\sigma^2(\theta) = \mathsf{D}_\theta \,\big(l(X_1, \theta_1) - l(X_1, 0)\big) \tag{1.6}$$

математическое ожидание и дисперсию случайных величин $l(X_1, \theta_1) - l(X_1, 0)$ при распределении P_θ.

Поскольку $\Lambda_n(\theta)$ есть сумма независимых одинаково распределенных случайных величин, то согласно центральной предельной теореме имеем при $n \to \infty$

$$\mathcal{L}\bigg(\frac{\Lambda_n(\theta_1) - n\mu(0)}{\sigma(0)\sqrt{n}} \;\Big|\; \mathsf{H}_0\bigg) \to \mathcal{N}(0, 1), \tag{1.7}$$

где $\mathcal{L}(Z \mid \mathsf{H}_i)$ означает распределение Z при гипотезе H_i, $i = 0, 1$, и $\mathcal{N}(\mu, \sigma^2)$ – нормальный закон с параметрами μ и σ^2. Из соотношения (1.4) теперь следует (поскольку сходимость функций распределения к нормальной функции распределения равномерна), что

$$c_\alpha(n) = \sqrt{n}\,\sigma(0)\,u_\alpha + n\mu(0) + o(1), \tag{1.8}$$

где $u_\alpha = \Phi^{-1}(1 - \alpha)$ и $\Phi(x)$ – функция распределения стандартного нормального закона.

Обозначим через $\beta_n^*(\theta_1)$ мощность наилучшего критерия для проверки гипотезы H_0 против альтернативы H_1 (см. (1.1), (1.2)), основанного на логарифме отношения правдоподобия $\Lambda_n(\theta_1)$, то есть

$$\beta_n^*(\theta_1) = \mathsf{P}_{n,\theta_1}\big(\Lambda_n(\theta_1) > c_\alpha(n)\big). \tag{1.9}$$

Покажем, что этот критерий состоятелен, то есть справедлив следующий хорошо известный результат (см., например, [6], теорема 3.3.1). Приведем здесь короткое доказательство этого факта, основанное на неравенстве Иенсена.

Лемма 1.1. *Если* $\sigma^2(0) > 0$, $\sigma^2(\theta_1) > 0$, *то*

$$\beta_n^*(\theta_1) \to 1, \quad n \to \infty.$$

Доказательство. Используя опять центральную предельную теорему, получаем

$$\mathcal{L}\left(\frac{\Lambda_n(\theta_1) - n\mu(\theta_1)}{\sigma(\theta_1)\sqrt{n}} \;\Big|\; \mathsf{H}_1\right) \to \mathcal{N}(0,1) \tag{1.10}$$

и поэтому с учетом (1.8) для мощности $\beta_n^*(\theta_1)$ имеем представление

$$\beta_n^*(\theta_1) = \mathsf{P}_{n,\theta_1}(\Lambda_n(\theta_1) > c_\alpha(n)) = 1 - \Phi\left(\frac{c_\alpha(n) - n\mu(\theta_1)}{\sigma(\theta_1)\sqrt{n}}\right) + o(1) =$$

$$= \Phi\left(\frac{\sqrt{n}(\mu(\theta_1) - \mu(0)) - u_\alpha\sigma(0)}{\sigma(\theta_1)}\right) + o(1). \tag{1.11}$$

Применим теперь неравенство Иенсена (см., например, [3], теорема 4.7.5) к математическим ожиданиям $\mu(\theta_1)$, $\mu(0)$ (см.(1.5)), имеем

$$\mu(0) = \mathsf{E}_0 \log\frac{p(X_1,\theta_1)}{p(X_1,0)} \le \log \mathsf{E}_0 \frac{p(X_1,\theta_1)}{p(X_1,0)} = 0,$$

$$\mu(\theta_1) = \mathsf{E}_{\theta_1} \log\frac{p(X_1,\theta_1)}{p(X_1,0)} = -\mathsf{E}_{\theta_1} \log\frac{p(X_1,0)}{p(X_1,\theta_1)} \ge -\log \mathsf{E}_{\theta_1} \frac{p(X_1,0)}{p(X_1,\theta_1)} = 0.$$

Из этих неравенств следует, что

$$\sqrt{n}\,\big(\mu(\theta_1) - \mu(0)\big) \to +\infty, \; n \to \infty,$$

и поэтому в силу формулы (1.11) действительно

$$\beta_n^*(\theta_1) \to 1, \quad n \to \infty.$$

□

Факт стремления мощности $\beta_n^*(\theta_1)$ к единице, а точнее скорость сходимости к единице, может быть использован для сравнения различных состоятельных критериев (см., например, [6] и обзор, приведенный там). Однако мы здесь рассмотрим несколько иной, асимптотический подход сравнения различных критериев. Это так называемый подход Питмэна (см. [7]), согласно которому для получения нетривиального предела мощности $\beta_n^*(\theta_1)$, заключенного между α и 1, рассматривают последовательность альтернатив $\theta_1 = \theta_n$, стремящуюся к нулю. Из соотношения (1.11) и центральной предельной теоремы для схемы серий (см., например, [3], теорема 8.4.5) следует, что в регулярном случае для выполнения этого должно быть

$$\mu(\theta_n) - \mu(0) = \mathcal{O}\big(n^{-1/2}\big), \quad \theta_n = \mathcal{O}\big(n^{-1/2}\big). \tag{1.12}$$

Поэтому далее будет рассматриваться задача проверки простой гипотезы H_0 (см. (1.1)) против последовательности сложных близких альтернатив вида

$$\mathsf{H}_{n,1}: \theta = \frac{t}{\sqrt{n}}, \quad 0 < t \leq C, \quad C > 0, \tag{1.13}$$

где параметр t неизвестен. Для любого фиксированного $t \in (0, C]$ наилучший критерий для проверки гипотезы H_0 против простой альтернативы

$$\mathsf{H}_{n,t}: \theta = \frac{t}{\sqrt{n}} \tag{1.14}$$

основан на логарифме отношения правдоподобия

$$\Lambda_n(t) \equiv \sum_{i=1}^{n} \big(l(X_i, tn^{-1/2}) - l(X_i, 0) \big). \tag{1.15}$$

Обозначим через $\beta_n^*(t)$ мощность такого критерия уровня $\alpha \in (0,1)$. Заметим, что поскольку t неизвестно, то нельзя использовать статистику $\Lambda_n(t)$ для построения критерия проверки гипотезы H_0 против альтернативы $\mathsf{H}_{n,1}$. Однако $\beta_n^*(t)$, это так называемая огибающая функций мощности, дает верхнюю границу для мощности любого критерия при проверке гипотезы H_0 против фиксированной альтернативы $\mathsf{H}_{n,t}$, $t > 0$, и может служить стандартом при сравнении различных критериев.

Найдем предельное выражение для $\beta_n^*(t)$. При естественных условиях регулярности формула Тейлора дает аппроксимацию

$$l\big(X_i, tn^{-1/2}\big) - l(X_i, 0) = \frac{t}{\sqrt{n}}\, l^{(1)}(X_i) + \frac{1}{2}\,\frac{t^2}{n}\, l^{(2)}(X_i) + \ldots, \tag{1.16}$$

где

$$l^{(j)}(x) = \frac{\partial^j}{\partial \theta^j}\, l(x,\theta)\Big|_{\theta = 0}, \quad j = 1, 2, \ldots .$$

Поэтому из (1.16) и (1.15) следует стохастическое разложение для логарифма отношения правдоподобия $\Lambda_n(t)$ вида

$$\Lambda_n(t) = tL_n^{(1)} - \frac{1}{2}t^2 I + \frac{1}{2}\,\frac{t^2}{\sqrt{n}}\, L_n^{(2)} + \ldots, \tag{1.17}$$

где

$$L_n^{(j)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(l^{(j)}(X_i) - \mathsf{E}_0\, l^{(j)}(X_i) \big), \quad j = 1, 2, \ldots,$$

и $I = \mathsf{E}_0\, (l^{(1)}(X_1))^2$ – фишеровская информация. При этом в выражении (1.17) опущены неслучайный член $\frac{1}{6}\frac{t^3}{\sqrt{n}}\mathsf{E}_0\, l^{(3)}(X_1)$, члены более высокого порядка малости, чем $n^{-1/2}$, и использован хорошо известный факт, состоящий в том, что

$$\mathsf{E}_0\, l^{(1)}(X_1) = 0, \quad \mathsf{E}_0\, l^{(2)}(X_1) = -I. \tag{1.18}$$

Критерий, основанный на статистике $\Lambda_n(t)$, отвергает гипотезу H_0 в пользу альтернативы $\mathsf{H}_{n,t}$, если

$$\Lambda_n(t) > c_{\alpha,t}(n), \tag{1.19}$$

где критическое значение $c_{\alpha,t}(n)$ выбирается из условия

$$\mathsf{P}_{n,0}\big(\Lambda_n(t) \;>\; c_{\alpha,t}(n)\big) \;=\; \alpha. \tag{1.20}$$

Аналогично из (1.7), (1.8) и (1.17) следует, что

$$\mathcal{L}\big(\Lambda_n(t) \mid \mathsf{H}_0\big) \;\to\; \mathcal{N}\Big(-\frac{1}{2}t^2 I, t^2 I\Big) \tag{1.21}$$

и

$$c_{\alpha,t}(n) \;\to\; c_{\alpha,t} \;=\; t\sqrt{I}\; u_\alpha \;-\; \frac{1}{2}t^2 I. \tag{1.22}$$

Найдем теперь предельное распределение $\Lambda_n(t)$ при альтернативе $\mathsf{H}_{n,t}$. Имеем (см. (1.17))

$$\mathsf{E}_{n,t/\sqrt{n}}\;\Lambda_n(t) \;=\; t\sqrt{n}\;\mathsf{E}_{t/\sqrt{n}}\;l^{(1)}(X_1) \;-\; \frac{1}{2}t^2 I \;+$$

$$+\;\frac{1}{2}t^2\;\Big(\mathsf{E}_{t/\sqrt{n}}\;l^{(2)}(X_1) \;-\; \mathsf{E}_0\;l^{(2)}(X_1)\Big) \;+\; \mathcal{O}\big(n^{-1/2}\big) \;=\; \frac{t^2}{2}I \;+\; \mathcal{O}\big(n^{-1/2}\big), \tag{1.23}$$

$$\mathsf{D}_{n,t/\sqrt{n}}\;\Lambda_n(t) \;=\; t^2\;\mathsf{D}_{n,t/\sqrt{n}}\;L_n^{(1)} \;+\; \mathcal{O}\big(n^{-1/2}\big) \;=\; t^2 I \;+\; \mathcal{O}\big(n^{-1/2}\big), \tag{1.24}$$

поэтому

$$\mathcal{L}(\Lambda_n(t) \mid \mathsf{H}_{n,t}) \;\to\; \mathcal{N}\Big(\frac{1}{2}t^2 I, t^2 I\Big). \tag{1.25}$$

Теперь с учетом (1.22), (1.25) и равенства

$$\beta_n^*(t) \;=\; \mathsf{P}_{n,t/\sqrt{n}}\big(\Lambda_n(t) \;>\; c_{\alpha,t}(n)\big)$$

имеем

$$\beta_n^*(t) \;\to\; \beta^*(t) \;=\; \Phi\big(t\sqrt{I} \;-\; u_\alpha\big). \tag{1.26}$$

Заметим, что для проверки гипотезы H_0 против альтернативы $\mathsf{H}_{n,1}$ существуют критерии, основанные на статистиках, отличных от $\Lambda_n(t)$, и имеющие ту же предельную мощность $\beta^*(t)$. Такие критерии называются асимптотически наиболее мощными (АНМ) (точнее, локально АНМ, поскольку альтернатива $\mathsf{H}_{n,1}$ имеет локальный характер). Таковы, например, критерии основанные на статистиках $L_n^{(1)}$, $\Lambda_n(t_0)$, где $t_0 \;>\; 0$ фиксировано, оценках максимального правдоподобия и т.п. Заметим, что все эти статистики не зависят от неизвестного параметра t, и поэтому могут быть использованы при проверке гипотезы H_0 против альтернативы $\mathsf{H}_{n,1}$.

Покажем, например, что критерий, основанный на статистике $L_n^{(1)}$ является АНМ и имеет предельную мощность $\beta^*(t)$. Имеем

$$L_n^{(1)} \;=\; \frac{1}{\sqrt{n}}\;\sum_{i=1}^{n}\;l^{(1)}(X_i),$$

$$\mathsf{E}_{n,0}\;L_n^{(1)} \;=\; \sqrt{n}\;\mathsf{E}_0\;l^{(1)}(X_1) \;=\; 0,$$

$$\mathsf{D}_{n,0}\;L_n^{(1)} \;=\; \mathsf{D}_0\;l^{(1)}(X_1) \;=\; I,$$

поэтому, если

$$\mathsf{P}_{n,0}\big(L_n^{(1)} \;>\; c_\alpha^{(1)}(n)\big) \;=\; \alpha \;\in\; (0,1),$$

то для критического уровня $c_\alpha^{(1)}(n)$ (в силу центральной предельной теоремы) имеем

$$c_\alpha^{(1)}(n) \;=\; \sqrt{I}\; u_\alpha \;+\; o(1). \tag{1.27}$$

Далее

$$\mathsf{E}_{n,t/\sqrt{n}}\; L_n^{(1)} \;=\; \sqrt{n}\; \mathsf{E}_{t/\sqrt{n}}\; l^{(1)}(X_1) \;=\; \sqrt{n}\int\; l^{(1)}(x)p\big(x,tn^{-1/2}\big)\; d\nu(x) \;=$$

$$=\; tI \;+\; \mathcal{O}\big(n^{-1/2}\big),$$

$$\mathsf{D}_{n,t/\sqrt{n}}\; L_n^{(1)} \;=\; \mathsf{D}_{t/\sqrt{n}}\; l^{(1)}(X_1) \;=\; \mathsf{E}_0\; (l^{(1)}(X_1))^2 \;+\; \mathcal{O}\big(n^{-1/2}\big) \;=\; I \;+\; \mathcal{O}\big(n^{-1/2}\big),$$

поэтому для мощности $\beta_n(t)$ критерия, основанного на статистике $L_n^{(1)}$, с учетом (1.27) имеем

$$\beta_n(t) \;=\; \mathsf{P}_{n,t/\sqrt{n}}\big(L_n^{(1)} \;>\; c_\alpha^{(1)(n)}\big) \;=\; \Phi\bigg(\frac{tI \;-\; c_\alpha^{(1)}(n)}{\sqrt{I}}\bigg) \;+\; o(1) \;=$$

$$=\; \Phi\big(t\sqrt{I} \;-\; u_\alpha\big) \;+\; o(1) \;\to\; \beta^*(t) \;=\; \Phi\big(t\sqrt{I} \;-\; u_\alpha\big). \tag{1.28}$$

Соотношение (1.26) создает естественную основу для асимптотического сравнения различных АНМ критериев, однако, для различения критериев такого рода, то есть удовлетворяющих соотношению

$$\beta_n(t) \;\to\; \beta^*(t), \tag{1.29}$$

где $\beta_n(t)$ – мощность конкретного рассматриваемого критерия, нужны следующие члены асимптотического разложения $\beta_n(t)$, то есть представления типа

$$\beta_n(t) \;=\; \beta^*(t) \;+\; \frac{1}{\sqrt{n}}h_1(t) \;+\; \frac{1}{n}h_2(t) \;+\; \cdots\,. \tag{1.30}$$

Асимптотическим разложениям в статистике посвящено много работ (см., например, работы [4], [9] и [11]). При получении формул типа (1.30) для различных критериев, было замечено, что при выполнении естественных условий регулярности для АНМ критериев совпадают и члены $h_1(t)$, различия наступают в членах порядка n^{-1}. Этим вопросам посвящены работы [1], [2], [4] и [11]. При этом величина

$$r(t) \;=\; \lim_{n\to\infty}\; n(\beta_n^*(t) \;-\; \beta_n(t)) \tag{1.31}$$

допускает статистическую интерпретацию в терминах необходимого числа наблюдений и позволяет находить асимптотический дефект (см., например, работы [8], [1], [2], [11], [17]).

Соотношение (1.30) может быть понято следующим образом. Предположим, что статистику T_n АНМ критерия можно монотонным преобразованием (не меняющим мощности критерия) преобразовать в статистику $S_n(t)$ такую, что величина

$$\Delta_n(t) \;\equiv\; S_n(t) \;-\; \Lambda_n(t) \;\to\; 0,\;\; n \;\to\; \infty \tag{1.32}$$

по вероятности относительно распределений $\mathsf{P}_{n,0}$ и $\mathsf{P}_{n,t/\sqrt{n}}$. Тогда критерий, основанный на статистике $S_n(t)$, имеет те же предельные распределения при гипотезах

H_0 и $\mathsf{H}_{n,t}$, что и критерий, основанный на $\Lambda_n(t)$, и, следовательно, ту же предельную мощность $\beta^*(t)$ (см. (1.26)). Например, в последнем примере

$$T_n \; = \; L_n^{(1)},$$

тогда, полагая (см. (1.17))

$$S_n(t) \; = \; tT_n \; - \; \frac{1}{2}t^2 I,$$

получим

$$\Delta_n(t) \; = \; -\frac{1}{2}\frac{t^2}{\sqrt{n}}L_n^{(2)} \; + \; \ldots \to \; 0. \tag{1.33}$$

В том типичном случае, когда $\Delta_n(t)$, как в (1.32), имеет порядок $n^{-1/2}$, то есть разность между $S_n(t)$ и $\Lambda_n(t)$ имеет тот же порядок, то можно ожидать, что мощность $\beta_n(t)$ критерия, основанного на $S_n(t)$ (или на T_n), отличается от $\beta_n^*(t)$ на величину порядка $n^{-1/2}$. Однако было обнаружено, что для широкого класса АНМ критериев это отличие имеет порядок n^{-1} (см. работы [2] и [1]). Первоначально выражения для $r(t)$ (см. (1.31)) строились с помощью асимптотических разложений для $\beta_n^*(t)$ и $\beta_n(t)$ (см. работы [4], [8] и [11]). Этот подход технически очень трудоемкий и громоздкий. Однако в работах [2] и [1] была получена общая формула для величины $r(t)$ без построения асимптотических разложений. Эта формула имеет наглядный вид в терминах условных дисперсий. Для ее демонстрации обозначим через $\Lambda(t)$ нормальную случайную величину вида $\mathcal{N}(-\frac{1}{2}t^2 I, t^2 I)$, тогда в силу (1.21)

$$\mathcal{L}(\Lambda_n(t) \mid \mathsf{H}_0) \; \to \; \mathcal{L}(\Lambda(t)), \tag{1.34}$$

и предположим, что при гипотезе H_0 случайный вектор

$$(\sqrt{n}\Delta_n(t), \; \Lambda_n(t)) \tag{1.35}$$

имеет предельное распределение (типичным образом двумерное нормальное), совпадающее с распределением вектора

$$(\Delta(t), \; \Lambda(t)), \tag{1.36}$$

тогда в работах [2] и [1] показано, что

$$r(t) \; = \; \frac{1}{2t\sqrt{I}}\varphi(u_\alpha \; - \; t\sqrt{I}) \; \mathsf{D}(\Delta(t) \mid \Lambda(t) \; = \; c_t), \tag{1.37}$$

где $c_t \; = \; t\sqrt{I}u_\alpha \; - \; \frac{1}{2}t^2 I$ и $\varphi(x) \; = \; \Phi'(x)$. Например, для критерия, основанного на статистике $T_n \; = \; L_n^{(1)}$ в работе [2] (см. там формулу (1.4.10)) получено выражение

$$r(t) \; = \; \frac{t^3}{8\sqrt{I}}\varphi(u_\alpha \; - \; t\sqrt{I})(\mathsf{D}_0 \; l^{(2)}(X_1) \; - \; I^{-1}\mathsf{E}_0^2 l^{(1)}(X_1) l^{(2)}(X_1)).$$

В работе [2] рассмотрен общий случай в терминах общего статистического эксперимента и приведена общая теорема ([2], теорема 3.2.1, см. также [10]), дающая достаточные условия для существования предела

$$r \; = \; \lim_{n\to\infty} \tau_n^{-2}(\beta_n^* \; - \; \beta_n) \; = \; \frac{1}{2}e^b p(b) \; \mathsf{D}(\Delta \mid \Lambda \; = \; b), \tag{1.38}$$

где $b = \Phi_1^{-1}(1 - \alpha)$, $\Phi_1(x)$ – функция распределения, предельная для логарифма отношения правдоподобия Λ_n при гипотезе H_0 (выше было $\Phi_1(x) = \Phi\Big(\frac{x+\frac{1}{2}t^2 I}{t\sqrt{I}}\Big)$), $p(x) = \Phi_1'(x)$ и $\tau_n \to 0$ – малый параметр (выше было $\tau_n = n^{-1/2}$), (Δ, Λ) – случайный вектор, предельный для $(\tau_n^{-1}\Delta_n, \Lambda_n)$, $\Delta_n = S_n - \Lambda_n$ и S_n – монотонное преобразование статистики критерия T_n, невлияющее на мощность соответствующего критерия.

Цель настоящей работы – рассмотрение и исследование асимптотического поведения обобщенных мощностей критериев, основанных на выборках случайного объема и получение для них асимптотических разложений (а.р.).

В работе приняты следующие обозначения: $\mathbb{R}$ – множество вещественных чисел, $\mathbb{N}$ – множество натуральных чисел, $\Phi(x)$, $\varphi(x)$ – соответственно ф.р. и плотность стандартного нормального закона.

2. Случай выборок случайного объема

В классических задачах математической статистики объем выборки или количество наблюдений, доступной исследователю, традиционно считается детерминированным и в асимптотических постановках играет роль (как правило, неограниченно возрастающего) *известного* параметра. Однако на практике часто возникают ситуации, когда размер выборки не является заранее определенным и может рассматриваться как случайный. Подобного рода ситуации, как правило, связаны с тем, что статистические данные накапливаются в течение фиксированного «времени». Это имеет место, в частности, в страховании, когда в течение разных отчетных периодов одинаковой длины (скажем, месяцев или лет) происходит разное число страховых событий (страховых выплат или заключений страховых контрактов), в медицине, когда число пациентов с тем или иным заболеванием варьируется от года к году, в технике, когда при испытании на надежность (скажем, при определении наработки на отказ) разных партий приборов (изделий), число отказавших приборов в разных партиях будет разным и заранее неопределенным. В таких ситуациях число наблюдений, которые будут доступны исследователю, и заранее не известное, разумно считать случайной величиной. Другими словами, в таких ситуациях объем выборки является не (известным) параметром, а сам становится *наблюдением*, то есть статистикой. В силу указанных обстоятельств вполне естественным становится изучение асимптотического поведения распределений статистик достаточно общего вида, основанных на выборках случайного объема.

На естественность такого подхода, в частности, обратил внимание Б.В. Гнеденко в работе [12], в которой рассматривались асимптотические свойства распределений выборочных квантилей, построенных по выборкам случайного объема, и было продемонстрировано, что при замене неслучайного объема выборки случайной величиной асимптотические свойства статистик могут радикально измениться. К примеру, если объем выборки является геометрически распределенной случайной величиной, то вместо ожидаемого в соответствии с классической теорией нормального закона, в качестве асимптотического распределения выборочной медианы возникает распределение Стьюдента с двумя степенями свободы, хвосты которого столь тяжелы, что у него отсутствуют моменты порядков, больших второго. «Тяжесть» хвостов асимптотических распределений же имеет критически важное значение, в частности, в задачах проверки гипотез.

Простейшей статистикой является сумма наблюдений. Для выборок случайного объема число слагаемых в таких суммах само становится случайным. Асимптотическим свойствам распределений сумм случайного числа случайных величин посвящено много работ (см., например, [12], [13], [15]). Такого рода суммы находят широкое применение в страховании, экономике, биологии и т.п. (см., [14], [15]). В классической статистике суммирование наблюдений как правило возникает при определении выборочных средних. При статистическом анализе, основанном на моделях, в которых объем выборки считается неслучайным, асимптотическое поведение статистик типа сумм и статистик типа средних арифметических одинаково – эти статистики после нормировки, обязательной для получения нетривиальных предельных распределений, становятся неразличимыми. Однако, как уже говорилось, в реальной практике очень часто объем выборки сам является статистикой, и, как недавно показано, например, в работе [16], асимптотическое поведение статистик типа сумм и статистик типа средних арифметических при их неслучайной нормировке оказывается различным. Заметим, что, конечно же, формально допустима и случайная нормировка, но для построения разумных асимптотических аппроксимаций для распределений статистик (а именно это и является целью асимптотической статистики), она неприменима. Именно использованием неслучайной нормировки и объясняется возникновение не «чистого» нормального закона, а смешанных нормальных предельных распределений у статистик типа сумм и типа средних арифметических. При этом различие этих предельных законов может дать дополнительную информацию о структуре исходных данных.

В задачах асимптотической проверки гипотез, как отмечено в Разделе 1, важной характеристикой качества критерия является его мощность и ее асимптотическая аппроксимация позволяет сравнивать асимптотически эквивалентные критерии. При этом, как показано ниже, у критериев, основанных на выборках случайного объема, типичным образом изменяется асимптотическое разложение обобщенной мощности в членах порядка n^{-1}, что приводит к конечности асимптотического дефекта (см. [1] и [8]).

3. Критерии, основанные на выборках случайного объема

В этом разделе будут рассмотрены критерии, основанные на выборках случайного объема. Итак, пусть $\mathbf{X}_n = (X_1, \ldots, X_n)$ – независимые одинаково распределенные наблюдения, каждое из которых принимает значения в произвольном измеримом пространстве $(\mathcal{X}, \mathcal{A})$ и имеет неизвестную с точностью до параметра θ плотность $p(x, \theta)$ относительно некоторой σ–конечной меры на $\mathcal{A}$. Предположим, что неизвестный параметр θ принадлежит открытому множеству $\Theta \subset \mathbb{R}^1$, содержащему ноль.

Предположим, что случайные величины (с.в.) $N_1,\ N_2, \ldots$ и $X_1,\ X_2, \ldots$, заданы на одном и том же вероятностном пространстве $(\Omega,\ \mathcal{B},\ \mathsf{P})$. Как указано выше, с.в. $X_1,\ X_2, \ldots\ X_n$ имеют смысл наблюдений, n – неслучайный объем выборки, а с.в. N_n – случайный объем выборки, зависящий от натурального параметра $n \in\ \mathbb{N}$.

Например, если с.в. N_n имеет геометрическое распределение вида

$$\mathsf{P}(N_n \;=\; k) \;=\; \frac{1}{n}\left(1 \;-\; \frac{1}{n}\right)^{k-1},\; k \in \mathbb{N},$$

то $\mathsf{E}\, N_n \;=\; n$, то есть среднее значение случайного объема выборки равно n.

Пусть для каждого $n \geq 1$ с.в. N_n принимает только натуральные значения (то есть, $N_n \in \mathbb{N}$) и при каждом $\theta \in \Theta$ не зависит от последовательности с.в. $X_1, X_2, \ldots$. Для любого $n \in \mathbb{N}$ обозначим через $T_n \;=\; T_n(X_1, \ldots, X_n)$ некоторую статистику, то есть действительную измеримую функцию, зависящую от наблюдений $X_1, \ldots, X_n$. Определим статистику T_{N_n}, зависящую от выборки случайного объема как

$$T_{N_n}(\omega) \;\equiv\; T_{N_n(\omega)}(X_1(\omega), \ldots, X_{N_n(\omega)}(\omega)), \quad \omega \;\in\; \Omega.$$

Всюду далее предполагается, что

$$\mathsf{E}\, N_n \;=\; n, \tag{3.1}$$

то есть средний размер выборки случайного объема равен размеру неслучайной выборки.

Рассмотрим задачу проверки простой гипотезы

$$\mathsf{H}_{n,0} : \theta = 0 \tag{3.2}$$

против последовательности простых альтернатив вида

$$\mathsf{H}_{n,1} : \theta \;=\; \theta_n \;\neq\; 0, \quad n \;\in\; \mathbb{N}, \tag{3.3}$$

основанную на выборке неслучайного объема $\mathbf{X}_n \;=\; (\, X_1, \ldots, X_n)$. Пусть $\Psi_n^*(\mathbf{X}_n)$ – некоторый критерий уровня $\alpha \;\in\; (0,1)$, то есть

$$\mathsf{E}_{n,0}\; \Psi_n^*(\mathbf{X}_n) \;=\; \alpha, \quad n \;\in\; \mathbb{N}$$

для проверки гипотезы (3.2) против последовательности альтернатив (3.3). Обозначим его мощность через

$$\beta_n^*(\theta_n) \;=\; \mathsf{E}_{n,\theta_n}\; \Psi_n^*(\mathbf{X}_n). \tag{3.4}$$

Предположим, что для мощности (3.4) справедлива следующая аппроксимация

$$\left|\beta_n^*(\theta_n) \;-\; \beta_\alpha \;-\; \sum_{i=1}^{m} \gamma_{n,i}\; \beta_{i,\alpha}\right| \;\leq\; C^*\; \delta_n, \quad n \;\in\; \mathbb{N}, \tag{3.5}$$

где $\beta_\alpha,\; \beta_{i,\alpha},\; i = 1, \ldots, m,\; m \in \mathbb{N},\; C^* > 0$ – некоторые числа и

$$\max_{1 \leq i \leq m} \gamma_{n,i} \;\to\; 0, \quad \delta_n \;\to\; 0, \quad n \;\to\; \infty.$$

Рассмотрим теперь выборку случайного объема N_n, то есть выборку $\mathbf{X}_{N_n} \;=\; (\, X_1, \ldots, X_{N_n})$ и при каждом $k \in \mathbb{N}$ на множестве

$$\{\omega \in \Omega : \mathbb{N}_n \;=\; k\}$$

рассмотрим задачу проверки гипотезы (3.2) против последовательности альтернатив (см. (3.3))

$$\mathsf{H}_{k,1} : \theta \;=\; \theta_k \;\neq\; 0. \tag{3.6}$$

Для этой задачи рассмотрим критерий вида

$$\Psi_n \;=\; \Psi^*_{N_n}(\mathbf{X}_{N_n}), \tag{3.7}$$

тогда он имеет обобщенный уровень значимости

$$\sum_{k=1}^{\infty} \mathsf{E}_{k,0}\ \Psi^*_k(\mathbf{X}_k)\ \mathsf{P}(N_n \;=\; k) = \;\alpha \tag{3.8}$$

и обобщенную мощность

$$\beta_n \;=\; \sum_{k=1}^{\infty} \mathsf{E}_{k,\theta_k}\ \Psi^*_k(\mathbf{X}_k)\ \mathsf{P}(N_n \;=\; k). \tag{3.9}$$

Лемма 3.1. *Пусть выполнено неравенство (3.5), тогда для обобщенной мощности β_n справедлива оценка*

$$\big|\beta_n \;-\; \beta_\alpha \;-\; \sum_{i=1}^{m} \beta_{i,\alpha}\ \mathsf{E}\ \gamma_{N_n,i}\big| \;\leq\; C^*\ \mathsf{E}\ \delta_n.$$

Доказательство непосредственно следует из определения обобщенной мощности (3.9) и неравенства (3.5).

Следствие 3.1. *Типичным образом (см. Раздел 1 и (1.13))*

$$\theta_n \;=\; \frac{t}{\sqrt{n}}, \qquad 0\ ,\ t \;\leq\; C, \quad C \;>\; 0,$$

и для мощности справедлива аппроксимация

$$\big|\beta^*_n(\frac{t}{\sqrt{n}}) \;-\; \beta_\alpha(t) \;-\; \frac{\beta_{1,\alpha}(t)}{\sqrt{n}} \;-\; \frac{\beta_{2,\alpha}(t)}{n}\big| \;\leq\; \frac{C_\alpha}{n^{1+\delta}},\;\; \delta \;>\; 0,\; C_\alpha \;>\; 0,$$

где $\beta_\alpha(t)$, $\beta_{1,\alpha}(t)$, $\beta_{2,\alpha}(t)$ – некоторые функции (в работе [4] приведены многочисленные примеры выполнения этого равенства). В этом случае утверждение Леммы означает, что

$$\big|\beta_n(t) \;-\; \beta_\alpha(t) \;-\; \beta_{1,\alpha}(t)\ \mathsf{E}\ N_n^{-1/2} \;-\; \beta_{2,\alpha}(t)\ \mathsf{E}\ N_n^{-1}\big| \;\leq\; C_\alpha\ \mathsf{E}\ N_n^{-1-\delta}.$$

Если при этом для случайного объема выборки N_n справедливы равенства

$$\mathsf{E}\ N_n^{-1/2} \;=\; \frac{a}{\sqrt{n}} \;+\; \frac{b}{n} \;+\; o(n^{-1}),\;\; a,\; b \;\in\; \mathbb{R},$$

$$\mathsf{E}\ N_n^{-1} \;=\; \frac{e}{n} \;+\; o(n^{-1}),\;\; e \;\in\; \mathbb{R},\;\; \mathsf{E}\ N_n^{-1-\delta} \;=\; o(n^{-1}),$$

то
$$\beta_n(t) \;=\; \beta_\alpha(t) \;+\; \frac{a\beta_{1,\alpha}(t)}{\sqrt{n}} \;+\; \frac{e\beta_{2,\alpha}(t) \;+\; b\beta_{1,\alpha}(t)}{n} \;+\; o(n^{-1})$$

и в случае, если $a = 1$ справедливо равенство (см.(1.31))

$$r(t) \;=\; \lim_{n\to\infty} n\big(\beta_n^*\big(t/\sqrt{n}\big) \;-\; \beta_n(t)\big) \;=\; (1 \;-\; e)\beta_{2,\alpha}(t) \;-\; b\beta_{1,\alpha}(t).$$

Приведем теперь пример применения Леммы 3.1 и ее следствия.

Пусть необходимо проверить простую гипотезу

$$\mathsf{H}_0 : \theta = 0 \tag{3.9}$$

на основе выборки $\mathbf{X}_n \;=\; (\,X_1,\ldots,X_n)$ против простой ($t \;\in\; [0,C],\ C \;>\; 0$ – фиксировано) альтернативы вида (см. формулу (1.13) и мотивацию естественности такого подхода перед ней)

$$\mathsf{H}_{n,t} : \theta \;=\; \frac{t}{\sqrt{n}}. \tag{3.10}$$

По лемме Неймана – Пирсона наилучший критерий уровня $\alpha \;\in\; (0,1)$ основан на логарифме отношения правдоподобия (см. Раздел 1)

$$\Lambda_n(t) \;\equiv\; \sum_{i=1}^{n} \big(l(X_i, tn^{-1/2}) \;-\; l(X_i,0)\big). \tag{3.11}$$

Обозначим, как и выше, через $\beta_n^*(t)$ мощность этого критерия, то есть

$$\psi_n^*(\mathbf{X}_n) \;=\; \begin{cases} 1, & \Lambda_n(t) \;>\; c_{\alpha,t}(n), \\ 0, & \Lambda_n(t) \;<\; c_{\alpha,t}(n), \end{cases} \tag{3.12}$$

причем
$$\mathsf{E}_{n,0}\ \psi_n^*(\mathbf{X}_n) \;=\; \alpha \tag{3.13}$$

и
$$\beta_n^*(t) \;=\; \mathsf{E}_{n,t/\sqrt{n}}\ \psi_n^*(\mathbf{X}_n). \tag{3.14}$$

Асимптотическое поведение этой мощности описывается следующей Теоремой.

Теорема 3.2. ([11], Лемма 6.1, стр. 136) *Пусть $p(x,\theta) \;=\; p(x \;-\; \theta)$, причем для любого $x \;\in\; \mathbb{R}$ справедливы соотношения $p(x) \;>\; 0,\;\; p(-x) \;=\; p(x)$, существует пятая производная $p^{(5)}(x)$ функции $p(x)$ и*

$$\limsup_{y\to 0} \int_{-\infty}^{+\infty} \Big|\frac{p^{(i)}(x+y)}{p(x+y)}\Big|^{5/i} p(x)\ dx \;<\; \infty, \quad i \;=\; 1,\ldots,5.$$

Тогда существует постоянная $A \;>\; 0$ такая, что справедливо неравенство

$$\sup_{t\in[0,C]} \big|\beta_n^*(t) \;-\; \bar\beta_n^*(t)\big| \;\le\; \frac{A}{n^{3/2}}, \tag{3.15}$$

где
$$\bar\beta_n^*(t) \;=\Phi(t\sqrt{I} \;-\; u_\alpha) \;+$$

$$+\ \frac{t\ \phi(t\sqrt{I}\ -\ u_\alpha)}{72n\ I^{3/2}}\ \Bigg(\mathsf{E}_0\ (l^{(1)}(X_1))^4\ \big(3(u_\alpha^2\ -\ 1)\ -\ 3t\sqrt{I}u_\alpha\ +\ 2t^2 I\big)\ -$$

$$-\ 3\mathsf{E}_0\ \left(\frac{p^{(2)}(X_1)}{p(X_1)}\right)^2\ t^2 I\ -\ 9I^2\big((u_\alpha^2\ -\ 1)\ -\ t\sqrt{I}u_\alpha\big)\Bigg).$$

Рассмотрим теперь критерий для проверки гипотезы (3.9) против последовательности близких альтернатив (3.10), основанный на выборке случайного объема $\mathbf{X}_{N_n}\ =\ (X_1,\ldots,X_{N_n})$, вида

$$\psi_n(\mathbf{X}_{N_n})\ \equiv\ \psi^*_{N_n}(\mathbf{X}_{N_n})\ =\ \begin{cases}1, & \Lambda_{N_n}(t)\ >\ c_{\alpha,t}(N_n),\\ 0, & \Lambda_{N_n}(t)\ <\ c_{\alpha,t}(N_n),\end{cases} \tag{3.16}$$

тогда (см. (3.13))

$$\mathsf{E}_{k,0}\ \psi_k(\mathbf{X}_k)\ =\ \alpha, \quad k\ \in\ \mathbb{N}$$

и обобщенный уровень значимости этого критерия есть

$$\sum_{k=1}^{\infty}\ \mathsf{E}_{k,0}\ \psi_k(\mathbf{X}_k)\ \mathsf{P}(N_n\ =\ k) =\ \alpha. \tag{3.17}$$

Обозначим обобщенную мощность этого критерия через

$$\beta_n(t)\ =\ \sum_{k=1}^{\infty}\ \mathsf{E}_{k,t/\sqrt{k}}\ \psi_k(\mathbf{X}_k)\ \mathsf{P}(N_n\ =\ k). \tag{3.18}$$

Применяя теперь Лемму 3.1 и ее Следствие, непосредственно получаем следующее утверждение.

Теорема 3.3. *Пусть выполнены условия Теоремы 3.2 и случайный индекс N_n удовлетворяет условиям Следствия 3.1. Тогда для обобщенной мощности $\beta_n(t)$ (см. (3.18)) критерия $\psi_n(\mathbf{X}_{N_n})$ (см. (3.16)), основанного на случайной выборке $\mathbf{X}_{N_n}\ =\ (X_1,\ldots,X_{N_n})$ справедливо a.p.*

$$\sup_{t\in[0,C]}\ \big|\beta_n(t)\ -\ \bar\beta_n(t)\big|\ =\ o(n^{-1}),$$

где

$$\bar\beta_n(t)\ = \Phi(t\sqrt{I}\ -\ u_\alpha)\ +$$

$$+\ \frac{e\ t\ \phi(t\sqrt{I}\ -\ u_\alpha)}{72n\ I^{3/2}}\ \Bigg(\mathsf{E}_0\ (l^{(1)}(X_1))^4\ \big(3(u_\alpha^2\ -\ 1)\ -\ 3t\sqrt{I}u_\alpha\ +\ 2t^2 I\big)\ -$$

$$-\ 3\mathsf{E}_0\ \left(\frac{p^{(2)}(X_1)}{p(X_1)}\right)^2\ t^2 I\ -\ 9I^2\big((u_\alpha^2\ -\ 1)\ -\ t\sqrt{I}u_\alpha\big)\Bigg)$$

и

$$r(t)\ =\ \lim_{n\to\infty} n\big(\beta^*_n(t)\ -\ \beta_n(t)\big)\ =$$

$$=\ \frac{(1\ -\ e)\ t\ \phi(t\sqrt{I}\ -\ u_\alpha)}{72\ I^{3/2}}\ \Bigg(\mathsf{E}_0\ (l^{(1)}(X_1))^4\ \big(3(u_\alpha^2\ -\ 1)\ -\ 3t\sqrt{I}u_\alpha\ +\ 2t^2 I\big)\ -$$

$$-\ 3\mathsf{E}_0\ \left(\frac{p^{(2)}(X_1)}{p(X_1)}\right)^2\ t^2 I\ -\ 9I^2\big((u_\alpha^2\ -\ 1)\ -\ t\sqrt{I}u_\alpha\big)\Bigg).$$

Замечание 1. Применяя результаты работы [11], аналогично этой Теореме можно получить асимптотические разложения для обобщенных мощностей ранговых критериев и критериев перестановок, основанных на выборках случайного объема, используемых в проблеме симметрии.

Применим теперь предыдущие результаты для получения а.р. обобщенной функции мощности $\beta_n(t)$ (см. (3.18)) критерия $\psi_n(\mathbf{X}_{N_n})$ (см. (3.16)) в случае, если объем выборки N_n имеет трехточечное симметричное распределения.

Пусть случайный индекс N_n (случайный объем выборки) имеет симметричное распределение вида

$$N_n \ : \ \begin{matrix} n \ - \ h_n, & n, & n \ + \ h_n \\ \frac{1}{3} & \frac{1}{3}, & \frac{1}{3}, \end{matrix} \tag{3.19}$$

где последовательность натуральных чисел $h_n \ < \ n$ удовлетворяет условию

$$\lim_{n\to\infty} \frac{h_n}{n} \ = \ 0. \tag{3.20}$$

Лемма 3.4. *Пусть случайная величина N_n имеет распределение (3.19) и выполнено условие (3.20), тогда справедливы равенства*

$$\mathsf{E}\ N_n \ = \ n,$$

$$\mathsf{E}\ N_n^{-1/2} \ = \ \frac{1}{\sqrt{n}} \ - \frac{1}{4\sqrt{n}} \left(\frac{h_n}{n}\right)^2 \ + \ O\Big(\frac{1}{\sqrt{n}} \left(\frac{h_n}{n}\right)^3\Big),$$

$$\mathsf{E}\ N_n^{-1} \ = \ \frac{1}{n} \ + \frac{2}{3n} \left(\frac{h_n}{n}\right)^2 \ + \ O\Big(\frac{1}{n} \left(\frac{h_n}{n}\right)^4\Big),$$

$$\mathsf{E}\ N_n^{-3/2} \ = \ \frac{1}{n^{3/2}} \ + \ O\Big(\frac{1}{n^{3/2}} \left(\frac{h_n}{n}\right)^2\Big), \quad n \ \to \ \infty.$$

Доказательство. Доказательство Леммы следует из равенств

$$\mathsf{E}\ N_n^{-1} \ = \ \frac{3n^2 \ - \ h_n^2}{3n(n^2 \ - \ h_n^2)} \ =$$

$$= \ \frac{1}{n} \left(1 \ - \ \frac{h_n^2}{3n}\right) \left(1 \ + \ \frac{h_n^2}{n^2} \ + \ O\Big(\frac{h_n^4}{n^4}\Big)\right) \ =$$

$$= \ \frac{1}{n} \ + \ \frac{2}{3n} \left(\frac{h_n}{n}\right)^2 \ + \ O\Big(\frac{1}{n} \left(\frac{h_n}{n}\right)^4\Big),$$

$$\mathsf{E}\ N_n^{-3/2} \ = \ \frac{1}{3n^{3/2}} \left(\frac{1}{(1 \ - \ h_n/n)^{3/2}} \ + \ 1 + \ \frac{1}{(1 \ + \ h_n/n)^{3/2}}\right) \ =$$

$$= \ \frac{1}{n^{3/2}} \ + \ O\Big(\frac{1}{n^{3/2}} \left(\frac{h_n}{n}\right)^2\Big).$$

Аналогично доказываются оставшиеся соотношения. □

Применяя Лемму 3.1, непосредственно получаем следующее утверждение.

Лемма 3.5. *Пусть выполнены условия Леммы 3.4 и Теоремы 3.3, тогда*

$$\bar{\beta}_n(t) = \Phi(t\sqrt{I} - u_\alpha) +$$

$$+ \frac{t\,\phi(t\sqrt{I} - u_\alpha)}{72n\, I^{3/2}} \Big(\mathsf{E}_0\, (l^{(1)}(X_1))^4 \, \big(3(u_\alpha^2 - 1) - 3t\sqrt{I}u_\alpha + 2t^2 I\big) -$$

$$- 3\mathsf{E}_0 \left(\frac{p^{(2)}(X_1)}{p(X_1)}\right)^2 t^2 I - 9I^2\big((u_\alpha^2 - 1) - t\sqrt{I}u_\alpha\big)\Big)$$

и

$$r(t) = \lim_{n\to\infty} n\big(\beta_n^*(t) - \beta_n(t)\big) = 0.$$

Заключение

В работе получены асимптотические разложения для обобщенных мощностей критериев, основанных на выборках случайного объема. Эти разложения дают возможность сравнивать критерии, имеющие одинаковую предельную мощность (асимптотически эффективные критерии). Показано, что в симметричном случае критерии, основанные на выборках случайного объема, асимптотически неразличимы от критериев, основанных на выборках неслучайного объема.

Список литературы

[1] Bening V.E. Asymptotic Theory of Testing Statistical Hypotheses: Efficient Statistics, Optimality, Power Loss, and Deficiency. Utrecht: VSP, 2000. 277 p.

[2] Чибисов Д.М. Вычисление дефекта асимптотически эффективных критериев // Теория вероятностей и ее применения. 1985. Т. 30, № 2. С. 269–288. https://doi.org/10.1137/1130037

[3] Боровков А.А. Теория вероятностей. М.: Эдиториал УРСС, 2003. 470 с.

[4] Pfanzagl J. Asymptotic expansions in parametric statistical theory // Developments in Statistics. Ed. by P.R. Krishnaiah. Vol. 3. New York: Academic Press, 1980. Pp. 1–97.

[5] Lehmann E.L., Romano J.P. Testing Statistical Hypotheses. Third edition. Springer, 2005. 784 p.

[6] Никитин Я.Ю. Асимптотическая эффективность непараметрических критериев. М.: Наука, 1995. 250 с.

[7] Pitman E.J.G. Lecture notes on nonparametric statistical inference. Lectures given for the University of North Carolina. Institute of Statistics, 1948.

[8] Hodges J.L., Lehmann E.L. Deficiency // The Annals of Mathematical Statistics. 1970. Vol. 41, № 5. Pp. 783–801.

[9] Bickel P.J. Edgeworth expansions in nonparametric statistics // The Annals of Mathematical Statistics. 1974. Vol. 2, № 1. Pp. 1–20.

[10] Bickel P.J., Chibisov D.M., Van Zwet W.R. On efficiency of first and second order // International Statistical Review. 1981. Vol. 49. Pp. 169–175.

[11] Albers W., Bickel P.J., Van Zwet W.R. Asymptotic expansions for the power of distribution free tests in the one – sample problem // Annals of Statistics. 1976. Vol. 4, № 1. Pp. 108–156.

[12] Гнеденко Б.В. Об оценке неизвестных параметров распределения при случайном числе независимых наблюдений // Труды Тбилисского Математического института. 1989. Т. 92. С. 146–150.

[13] Гнеденко Б.В., Фахим Х. Об одной теореме переноса // Доклады Академии наук СССР. 1969. Т. 187. С. 15–17.

[14] Бенинг В.Е., Королев В.Ю. Об использовании распределения Стьюдента в задачах теории вероятностей и математической статистики // Теория вероятностей и ее применения. 2004. Т. 49, № 3. С. 417–435. https://doi.org/10.1137/S0040585X97981159

[15] Bening V.E., Korolev V.Yu. Generalized Poisson Models and Their Applications in Insurance and Finance. Utrecht: VSP, 2002. 434 p.

[16] Бенинг В.Е., Королев В.Ю. Некоторые статистические задачи, связанные с распределением Лапласа // Информатика и ее применения. 2008. Т. 2, № 2. С. 19–34.

[17] Бенинг В.Е. О дефекте некоторых оценок, основанных на выборках случайного объёма // Вестник ТвГУ. Серия: Прикладная математика. 2015. № 1. С. 5–14.

Образец цитирования

Бенинг В.Е. О мощности критериев в случае выборок случайного объема // Вестник ТвГУ. Серия: Прикладная математика. 2018. № 4. С. 5–22. https://doi.org/10.26456/vtpmk514

Сведения об авторах

1. **Бенинг Владимир Евгеньевич**
профессор кафедры математической статистики факультета вычислительной математики и кибернетики Московского государственного университета имени М.В. Ломоносова; старший научный сотрудник ИПИ РАН.

Россия, 119992, г. Москва, ГСП-1, Воробьевы горы, МГУ им. М.В. Ломоносова. E-mail: bening@yandex.ru

ON THE POWER OF CRITERIA IN THE CASE OF SAMPLES OF RANDOM SIZE

Bening Vladimir Evgenyevich
Professor at Mathematical Statistics department,
Lomonosov Moscow State University
Senior Researcher, Institute of Informatics Problems
of the Russian Academy of Sciences
Russia, 119992, Moscow, GSP-1, Vorobyovi gory, Lomonosov MSU.
E-mail: bening@yandex.ru

Received 01.11.2018, revised 03.12.2018.

In the paper general theorem concerning the asymptotic expansions for the generalized power of the tests based on the sample with random size is proved. These results are closed connected with the concept of asymptotic deficiency and may be used for asymptotic comparison of different asymptotically equivalent tests. Some examples are presented.

Keywords: power of test, asymptotic deficiency, sample of random size, asymptotic expansions, three-point symmetric distribution.

Citation

Bening V.E., "On the power of criteria in the case of samples of random size", *Vestnik TvGU. Seriya: Prikladnaya Matematika* [Herald of Tver State University. Series: Applied Mathematics], 2018, no. 4, 5–22. (in Russian) https://doi.org/10.26456/vtpmk514

References

[1] Bening V.E., *Asymptotic Theory of Testing Statistical Hypotheses: Efficient Statistics, Optimality, Power Loss, and Deficiency*, VSP, Utrecht, 2000, 277 pp.

[2] Chibisov D.M., "Calculation of deficiency of asymptotically efficient tests", *Theory of Probability and its Applications*, **30**:2 (1986), 289–310, https://doi.org/10.1137/1130037.

[3] Borovkov A.A., *Teoriya Veroyatnostej [Probability Theory]*, Editorial URSS Publ., Moscow, 2003 (in Russian), 470 pp.

[4] Pfanzagl J., "Asymptotic expansions in parametric statistical theory", *Developments in Statistics.* V. 3, ed. P.R. Krishnaiah, Academic Press, New York, 1980, 1–97.

[5] Lehmann E.L., Romano J.P., *Testing Statistical Hypotheses*, Third edition, Springer, 2005, 784 pp.

[6] Nikitin Ya.Yu., *Asimptoticheskaya Effektivnost Neparametricheskih Kriteriev [Asymptotic Efficiency of Non-Parametric Criteria]*, Nauka Publ., Moscow, 1995 (in Russian), 250 pp.

[7] Pitman E.J.G., *Lecture notes on nonparametric statistical inference*, Lectures given for the University of North Carolina, Institute of Statistics, 1948.

[8] Hodges J.L., Lehmann E.L., "Deficiency", *The Annals of Mathematical Statistics*, **41**:5 (1970), 783–801.

[9] Bickel P.J., "Edgeworth expansions in nonparametric statistics", *The Annals of Mathematical Statistics*, **2**:1 (1974), 1–20.

[10] Bickel P.J., Chibisov D.M., Van Zwet W.R., "On efficiency of first and second order", *International Statistical Review*, **49** (1981), 169–175.

[11] Albers W., Bickel P.J., Van Zwet W.R., "Asymptotic expansions for the power of distribution free tests in the one – sample problem", *Annals of Statistics*, **4**:1 (1976), 108–156.

[12] Gnedenko B.V., "On the estimation of unknown distribution parameters for a random number of independent observations", *Trudy Tbilisskogo Matematicheskogo instituta [Proceedings of the Tbilisi Mathematical Institute]*, **92** (1989), 146–150 (in Russian).

[13] Gnedenko B.V., Fakhim Kh., "On a transfer theorem", *Doklady Mathematics*, **187** (1969), 15–17 (in Russian).

[14] Bening V.E., Korolev V.Y., "On an application of the Student distribution in the theory of probability and mathematical statistics", *Theory of Probability and its Applications*, **49**:3 (2005), 377–391, https://doi.org/10.1137/S0040585X97981159.

[15] Bening V.E., Korolev V.Yu., *Generalized Poisson Models and Their Applications in Insurance and Finance*, VSP, Utrecht, 2002, 434 pp.

[16] Bening V.E., Korolev V.Yu., "Some statistical problems related to the Laplace distribution", *Informatics and Applications*, **2**:2 (2008), 19–34 (in Russian).

[17] Bening V.E., "On deficiencies of some estimators based on samples of random size", *Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics]*, 2015, № 1, 5–14 (in Russian).