ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
2017 Управление, вычислительная техника и информатика № 41

УДК 519.24
DOI: 10.17223/19988605/41/3

Б.Ю. Лемешко, С.Б. Лемешко, И.В. Веретельникова

О ПРИМЕНЕНИИ КРИТЕРИЕВ ПРОВЕРКИ ОДНОРОДНОСТИ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ

Работа выполнена при поддержке Министерства образования и науки РФ в рамках государственной работы «Обеспечение проведения научных исследований» (№ 1.4574.2017/6.7) и проектной части государственного задания (№ 1.1009.2017/4.6).

Проведен сравнительный анализ мощности критериев однородности законов распределения вероятностей (критериев Смирнова, Лемана–Розенблатта, Андерсона–Дарлинга). Построены модели предельных распределений статистик для *k*-выборочного критерия Андерсона–Дарлинга. Даются рекомендации по применению критериев.
Ключевые слова: критерии однородности; критерий Смирнова; критерий Лемана–Розенблатта; критерий Андерсона–Дарлинга; мощность критерия.

С необходимостью решения задач проверки гипотез о принадлежности двух (или более) выборок случайных величин одной и той же генеральной совокупности (проверки однородности) постоянно сталкиваются при анализе случайных ошибок средств измерений, при статистическом управлении качеством процессов. Такая задача естественно возникает при поверке средств измерений, когда пытаются убедиться в том, что закон распределения случайных ошибок измерений не претерпел существенных изменений по истечении некоторого интервала времени. При обработке результатов экспериментальных исследований такую задачу часто приходится решать технологам, медикам, биологам.

Задача проверки однородности двух выборок формулируется следующим образом. Пусть имеются две упорядоченные по возрастанию выборки размером m и n:

$$x_1 < x_2 < \ldots < x_m \quad \text{и} \quad y_1 < y_2 < \ldots < y_n .$$

Для определенности обычно полагают, что $m \le n$, но это совсем необязательно. Проверяется гипотеза о том, что две выборки извлечены из одной и той же генеральной совокупности, т.е. $H_0: \; F(x) = G(x)$ при любом x.

Как правило, на практике используется либо критерий Смирнова [1], либо критерий Лемана–Розенблатта [1–3]. В русскоязычной литературе практически не упоминается о применении критерия однородности Андерсона–Дарлинга [4] (Андерсона–Дарлинга–Петита) или, тем более, об использовании многовыборочного варианта этого критерия [5].

Настоящая статья является развитием работы [6], в которой были исследованы реальные свойства распределений статистик и мощности критериев Смирнова и Лемана–Розенблатта при ограниченных объемах выборок. В данном случае исследования дополнены анализом критериев Андерсона–Дарлинга и развитием его многовыборочного варианта. При проведении исследований использовалась методика компьютерного моделирования и исследования статистических закономерностей [7], хорошо зарекомендовавшая себя при сравнительном анализе критериев, ориентированных на проверку гипотез определённого вида [8–10].

1. Критерий Смирнова

Критерий однородности Смирнова предложен в работе [11]. Предполагается, что функции распределения $F(x)$ и $G(x)$ являются непрерывными. Статистика критерия Смирнова измеряет расстояние между эмпирическими функциями распределения, построенными по выборкам

$$D_{m,n} = \sup_x \left| G_m(x) - F_n(x) \right| .$$

При практическом использовании критерия значение статистики $D_{m,n}$ рекомендуется вычислять в соответствии с соотношениями [1]:

$$D_{m,n}^{+} = \max_{1 \le r \le m}\left[\frac{r}{m} - F_n(x_r)\right] = \max_{1 \le s \le n}\left[G_m(y_s) - \frac{s-1}{n}\right],$$

$$D_{m,n}^{-} = \max_{1 \le r \le m}\left[F_n(x_r) - \frac{r-1}{m}\right] = \max_{1 \le s \le n}\left[\frac{s}{n} - G_m(y_s)\right],$$

$$D_{m,n} = \max\left(D_{m,n}^{+}, D_{m,n}^{-}\right).$$

Для объёмов выборок $m, n\ \le 20$ таблицы процентных точек для статистики $D_{m,n}$ приводятся в [1]. Если гипотеза H_0 справедлива, то при неограниченном увеличении объемов выборок статистика

$$S_{\text{C}} = \sqrt{\frac{mn}{m+n}} D_{m,n} \qquad (1)$$

в пределе подчиняется распределению Колмогорова $K(S)$ [1].

Недостатки критерия Смирнова заключаются в следующем. Во-первых, при ограниченных значениях m и n случайные величины $D_{m,n}^{+}$ и $D_{m,n}$ являются дискретными, и множество их возможных значений представляет собой решетку с шагом $1/k$, где k – наименьшее общее, кратное m и n [1]. Гладкость распределения статистики зависит от величины k. Поэтому предпочтительнее применять критерий, когда объемы выборок m и n не равны и представляют собой взаимно простые числа. При равных объёмах выборок явная ступенчатость $G\left(S_{\text{C}} \middle| H_0\right)$ сохраняется даже при $m = n = 1000$.

Другим недостатком применения критерия со статистикой (1) является то, что распределения $G\left(S_{\text{C}} \middle| H_0\right)$ с ростом m и n приближаются к предельному распределению $K(s)$ слева. И при небольших и умеренных значениях m и n распределения $G\left(S_{\text{C}} \middle| H_0\right)$ существенно сдвинуто влево от $K(S)$. В этой связи в [6] предложена простая модификация статистики (1).

2. Критерий Лемана–Розенблатта

Критерий однородности Лемана–Розенблатта представляет собой критерий типа ω^2. Критерий предложен в работе [2] и исследован в [3]. Статистика критерия имеет вид [1]:

$$T = \frac{mn}{m+n} \int_{-\infty}^{\infty} \left[G_m(x) - F_n(x)\right]^2 dH_{m+n}(x),$$

где $H_{m+n}(x) = \frac{m}{m+n} G_m(x) + \frac{n}{m+n} F_n(x)$ – эмпирическая функция распределения, построенная по вариационному ряду объединения двух выборок. Статистика T используется в форме [Там же]:

$$T = \frac{1}{mn(m+n)}\left[n \sum_{i=1}^{n} \left(r_i - i\right)^2 + m \sum_{i=1}^{m} \left(s_i - j\right)^2\right] - \frac{4mn-1}{6(m+n)}, \qquad (2)$$

где r_i – порядковый номер (ранг) y_i; s_j – порядковый номер (ранг) x_j в объединенном вариационном ряде.

В [3] было показано, что статистика (2) в пределе распределена как $a1(t)$ [1]. В отличие от критерия Смирнова распределение статистики T быстро сходится к предельному $a1(T)$ [Там же]. При $m = n = 100$ распределение $G\left(T \middle| H_0\right)$ статистики (2) практически совпадает с $a1(T)$. При $m, n \le 25$ желательно учитывать отклонение реального распределения $G\left(T \middle| H_0\right)$ статистики (2) от $a1(T)$.

3. Критерий Андерсона–Дарлинга

Двухвыборочный критерий Андерсона–Дарлинга (критерий однородности) рассмотрен в работе [4]. Статистика критерия определяется выражением

$$A^2 = \frac{mn}{m+n}\int_{-\infty}^{\infty}\frac{[G_m(x)-F_n(x)]^2}{(1-H_{m+n}(x))H_{m+n}(x)}dH_{m+n}(x).$$

Для выборок непрерывных случайных величин выражение для этой статистики принимает простой вид [4]:

$$A^2 = \frac{1}{mn}\sum_{i=1}^{m+n-1}\frac{\left(M_i(m+n)-mi\right)^2}{i(m+n-i)}, \tag{3}$$

где M_i – число элементов первой выборки, меньших или равных i-му элементу вариационного ряда объединенной выборки.

Предельным распределением статистики (3) при справедливости проверяемой гипотезы H_0 является то же самое распределение $a2(t)$ [4], которое является предельным для статистики критерия согласия Андерсона–Дарлинга [1].

Вопросы мощности критерия рассматривались в [4, 12], мощность критерия исследовалась в [13]. Сходимость распределения $G\left(A^2 \middle| H_0\right)$ статистики (7) к $a2(A^2)$ при ограниченных объёмах выборок была исследована в [13], где было показано, что при $m, n \geq 45$ отклонение функции распределения $G\left(A^2 \middle| H_0\right)$ от $a2(A^2)$ не превышает 0,01. При $m = n = 100$ распределение $G\left(A^2 \middle| H_0\right)$ статистики (7) практически совпадает с $a2(A^2)$.

4. Анализ мощности двухвыборочных критериев

Мощность критериев проверки однородности исследовалась в случае различных альтернатив. В данном случае (для определенности) проверяемой гипотезе H_0 соответствовала принадлежность выборок одному и тому же стандартному нормальному закону распределения с плотностью

$$f(x) = \frac{1}{\theta_1\sqrt{2\pi}}\exp\left\{-\frac{(x-\theta_0)^2}{2\theta_1^2}\right\}$$

и параметрами сдвига $\theta_0 = 0$ и масштаба $\theta_1 = 1$.

При всех альтернативах первая выборка всегда соответствовала стандартному нормальному закону, а вторая – некоторому другому.

При альтернативе сдвига и конкурирующей гипотезе H_1 вторая выборка соответствовала нормальному закону с параметром сдвига $\theta_0 = 0{,}1$ и параметром масштаба $\theta_1 = 1$.

При изменении масштаба и конкурирующей гипотезе H_2 вторая выборка соответствовала нормальному закону с параметрами $\theta_0 = 0$ и $\theta_1 = 1{,}1$.

В случае конкурирующей гипотезы H_3 вторая выборка соответствовала логистическому закону с плотностью

$$f(x) = \frac{1}{\theta_1\sqrt{3}}\exp\left\{-\frac{\pi(x-\theta_0)}{\theta_1\sqrt{3}}\right\}\Bigg/\left[1+\exp\left\{-\frac{\pi(x-\theta_0)}{\theta_1\sqrt{3}}\right\}\right]^2$$

и параметрами $\theta_0 = 0$ и $\theta_1 = 1$. Нормальный и логистический законы очень близки и трудноразличимы с помощью критериев согласия.

Поскольку при ограниченных объемах выборок распределения статистик критериев существенно отличаются от предельных, оценки мощности находились по результатам моделирования распределений статистик при справедливости проверяемой H_0 и конкурирующих гипотез H_i, $i = \overline{1,3}$, при конкретных объемах выборок n. Количество экспериментов имитационного моделирования в каждом случае составило величину $N = 10^6$. Некоторые оценки мощности критериев Смирнова (С), Лемана–Розенблатта (ЛР), Андерсона–Дарлинга (АД) при заданных уровнях значимости $\alpha = 0{,}1$ и различных объемах выборок представлены в табл. 1.

Таблица 1

Оценки мощности критериев однородности относительно альтернатив $H_1 - H_3$ в зависимости от объемов выборок ($m = n$, $\alpha = 0{,}1$)

Критерий	$n = 20$	$n = 50$	$n = 100$	$n = 300$	$n = 500$	$n = 1000$	$n = 2000$
Относительно альтернативы H_1							
АД	0,114	0,137	0,175	0,319	0,447	0,691	0,919
ЛР	0,115	0,136	0,173	0,313	0,438	0,678	0,910
С	0,111	0,132	0,164	0,280	0,381	0,617	0,869
Относительно альтернативы H_2							
АД	0,104	0,112	0,128	0,202	0,290	0,528	0,861
ЛР	0,103	0,107	0,114	0,149	0,191	0,324	0,624
С	0,105	0,108	0,120	0,150	0,186	0,297	0,551
Относительно альтернативы H_3							
АД	0,103	0,108	0,117	0,156	0,203	0,343	0,640
С	0,104	0,110	0,121	0,159	0,198	0,319	0,564
ЛР	0,103	0,106	0,113	0,142	0,178	0,288	0,547

Критерий Андерсона–Дарлинга, как правило, обладает несколько большей мощностью, чем критерий Лемана–Розенблатта, особенно в случае отличия анализируемых выборок в характеристиках рассеяния. В то же время эксперименты показали, что относительно (очень) близких конкурирующих гипотез при малых объемах выборок преимущество в мощности может быть за критерием Лемана–Розенблатта. Критерий Смирнова заметно уступает в мощности критериям Лемана–Розенблатта и Андерсона–Дарлинга. В то же время следует обратить внимание на неплохие показатели критерия Смирнова относительно гипотезы H_3.

5. Многовыборочный критерий Андерсона–Дарлинга

Вопросы построения k-выборочных критериев однородности законов, являющихся аналогами критериев однородности Смирнова и Лемана–Розенблатта, рассматривались в работе [14]. Однако нам не известно о соответствующих результатах построения, которые можно было бы рекомендовать к применению.

Задача проверки однородности k выборок формулируется следующим образом. Пусть x_{ij} – j-е наблюдение i-й выборки $j = \overline{1, n_i}$, $i = \overline{1, k}$. Предположим, что i-й выборке соответствует непрерывная функция распределения $F_i(x)$. Необходимо проверить гипотезу вида $H_0 : F_1(x) = F_2(x) = \ldots = F_k(x)$ без указания общего для них закона распределения.

Многовыборочный вариант критерия однородности Андерсона–Дарлинга предложен в [5]. Обозначим эмпирическую функцию распределения, соответствующую i-й выборке, как $F_{in_i}(x)$, а эмпирическую функцию распределения, соответствующую объединенной выборке объемом $n = \sum_{i=1}^{k} n_i$, – как $H_n(x)$. Статистика k-выборочного критерия Андерсона–Дарлинга определяется выражением

$$A_{kn}^2 = \sum_{i=1}^{k} n_i \int_{B_n} \frac{[F_{in_i}(x) - H_n(x)]^2}{\big(1 - H_n(x)\big) H_n(x)} dH_n(x), \tag{4}$$

где $B_n = \{x \in R : H_n(x) < 1\}$. Для $k = 2$ соотношение (4) сводится к (3). В предположении о непрерывности $F_i(x)$, упорядочив объединенную выборку $Z_1 \le Z_2 \le \ldots \le Z_n$, непосредственно из (4) можно получить простое выражение для вычисления статистики

$$A_{kn}^2 = \frac{1}{n} \sum_{i=1}^{k} \frac{1}{n_i} \sum_{j=1}^{n-1} \frac{\left(nM_{ij} - jn_i\right)^2}{j(n-j)}, \tag{5}$$

где M_{ij} – число элементов в i-й выборке, которые не больше, чем Z_j. Проверяемая гипотеза H_0 отклоняется при больших значениях статистики (5).

В работе [5] таблица верхних процентных точек представлена не для статистики (5), а для статистики вида

$$T_{kn} = \frac{A_{kn}^2 - (k-1)}{\sqrt{D[A_{kn}^2]}}. \tag{6}$$

Дисперсия статистики A_{kn}^2 определяется выражением [5]:

$$D\left[A_{kn}^2\right] = \frac{an^3 + bn^2 + cn + d}{(n-1)(n-2)(n-3)}$$

при

$$a = (4g-6)(k-1) + (10-6g)H, \; b = (2g-4)k^2 + 8hk + (2g-14h-4)H - 8h + 4g - 6,$$
$$c = (6h+2g-2)k^2 + (4h-4g+6)k + (2h-6)H + 4h, \; d = (2h+6)k^2 - 4hk,$$

где

$$H = \sum_{i=1}^{k} \frac{1}{n_i}, \quad h = \sum_{i=1}^{n-1} \frac{1}{i}, \quad g = \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \frac{1}{(n-i)j}.$$

Зависимость предельных распределений статистики (6) от числа сравниваемых выборок k иллюстрирует рис. 1. С ростом числа сравниваемых выборок это распределение медленно сходится к стандартному нормальному закону. Исследование распределений статистик методами статистического моделирования при конкретных k показало наличие соответствующих предельных распределений. Результаты моделирования выявили, что при использовании критерия отличием распределений статистик от соответствующих предельных можно пренебречь при объемах анализируемых выборок $n_i \geq 30$.

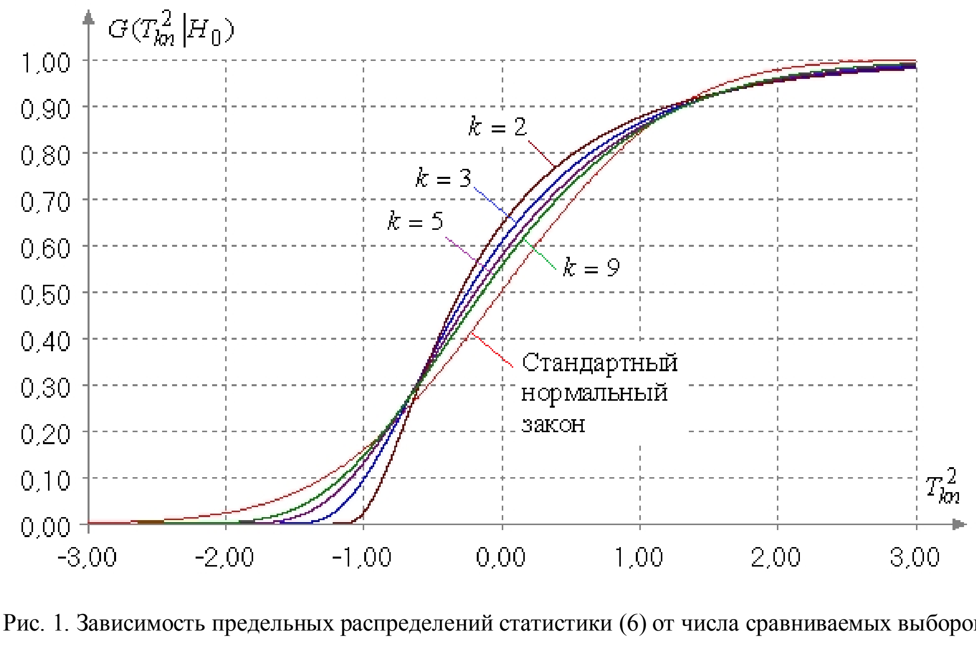

Рис. 1. Зависимость предельных распределений статистики (6) от числа сравниваемых выборок

Верхние критические значения предельных распределений для статистики (6), полученные методами статистического моделирования (при $n_i = 1000$ и числе экспериментов имитационного моделирования $N = 10^6$), уточняющие и расширяющие таблицу, приведенную в [5], представлены в [15]. В то же время для числа сравниваемых выборок $k = 2 \div 11$ нами были построены приближенные модели предельных распределений статистики (6). Хорошими моделями оказались законы семейства бета-распределений III рода с плотностью

$$f(x)=\frac{\theta_2^{\theta_0}}{\theta_3 \mathrm{B}(\theta_0,\theta_1)} \frac{\left(\frac{x-\theta_4}{\theta_3}\right)^{\theta_0-1}\left(1-\frac{x-\theta_4}{\theta_3}\right)^{\theta_1-1}}{\left[1+(\theta_2-1)\frac{x-\theta_4}{\theta_3}\right]^{\theta_0+\theta_1}}$$

при конкретных значениях параметров этого закона, найденных по полученным в результате моделирования выборкам статистик. Представленные в табл. 2 модели $\mathrm{B}_{\mathrm{III}}(\theta_0,\theta_1,\theta_2,\theta_3,\theta_4)$ с приведенными значениями параметров позволяют по значениям статистики, вычисленным по соотношению (6), находить оценки достигнутого уровня значимости p_{value} при соответствующем числе k сравниваемых выборок.

Таблица 2

Модели предельных распределений статистики

k	Модель
2	$\mathrm{B}_{\mathrm{III}}$ (3,1575; 2,8730; 18,1238; 15,0000; –1,1600)
3	$\mathrm{B}_{\mathrm{III}}$ (3,5907; 4,5984; 7,8040; 14,1310; –1,5000)
4	$\mathrm{B}_{\mathrm{III}}$ (4,2657; 5,7035; 5,3533; 12,8243; –1,7500)
5	$\mathrm{B}_{\mathrm{III}}$ (6,2992; 6,5558; 5,6833; 13,010; –2,0640)
6	$\mathrm{B}_{\mathrm{III}}$ (6,7446; 7,1047; 5,0450; 12,8562; –2,2000)
7	$\mathrm{B}_{\mathrm{III}}$ (6,7615; 7,4823; 4,0083; 11,800; –2,3150)
8	$\mathrm{B}_{\mathrm{III}}$ (5,8057; 7,8755; 2,9244; 10,900; –2,3100)
9	$\mathrm{B}_{\mathrm{III}}$ (9,0736; 7,4112; 4,1072; 10,800; –2,6310)
10	$\mathrm{B}_{\mathrm{III}}$ (10,2571; 7,9758; 4,1383; 11,186; –2,7988)
11	$\mathrm{B}_{\mathrm{III}}$ (10,6848; 7,5950; 4,2041; 10,734; –2,8400)
∞	N(0,0; 1,0)

В табл. 3 приведены оценки мощности k-выборочного критерия Андерсона–Дарлинга при $k=4$ относительно H_1, H_2 и H_3, когда только одна из k выборок принадлежала конкурирующему закону. При $k=2$ критерий со статистикой (6) эквивалентен по мощности двухвыборочному критерию Андерсона–Дарлинга со статистикой (3).

Естественно, что с ростом количества сравниваемых выборок тех же объемов мощность критерия относительно аналогичных конкурирующих гипотез снижается.

Таблица 3

Оценки мощности k-выборочного критерия однородности Андерсона–Дарлинга относительно альтернатив H_1 и H_3 в зависимости от объемов выборок (k = 4, $n_i = n$)

Уровень значимости α	n = 20	n = 50	n = 100	n = 300	n = 500	n = 1000
Относительно альтернативы H_1						
0,1	0,112	0,131	0,164	0,301	0,433	0,701
Относительно альтернативы H_2						
0,1	0,104	0,110	0,123	0,180	0,254	0,474
Относительно альтернативы H_3						
0,1	0,102	0,106	0,113	0,143	0,179	0,291

Заключение

Построенные в данной работе модели предельных распределений статистики (6) при использовании k-выборочного критерия однородности Андерсона–Дарлинга для анализа k сравниваемых выборок ($k=2\div 11$) дают возможность находить оценки p_{value}, что, несомненно, делает результаты статистических выводов более информативными и более обоснованными.

Необходимо упомянуть о критериях однородности Жанга [16, 17], которые дают возможность анализировать $k \geq 2$ выборок. Три критерия Жанга со статистиками Z_K , Z_C и Z_A являются развитием, соответственно, критериев однородности Смирнова, Лемана–Розенблатта и Андерсона–Дарлинга. Предварительные исследования показали, что критерии однородности Жанга имеют некоторое преимущество в мощности по отношению к альтернативам масштаба, но несколько уступают рассмотренным критериям Смирнова, Лемана–Розенблатта и Андерсона–Дарлинга по отношению к альтернативам сдвига. Недостатком критериев Жанга, затрудняющим их применение в приложениях, является зависимость распределений статистик от объемов выборок. Однако при современном уровне развития информационных технологий такой недостаток уже не является критичным [15].

ЛИТЕРАТУРА

1. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М. : Наука, 1983.
2. Lehmann E.L. Consistency and unbiasedness of certain nonparametric tests // Ann. Math. Statist. 1951. V. 22, No. 1. P. 165–179.
3. Rosenblatt M. Limit theorems associated with variants of the von Mises statistic // Ann. Math. Statist. 1952. V. 23. P. 617–623.
4. Pettitt A.N. A two-sample Anderson-Darling rank statistic // Biometrika. 1976. V. 63, No. 1. P. 161–168.
5. Scholz F.W., Stephens M.A. K-Sample Anderson–Darling Tests // Journal of the American Statistical Association. 1987. V. 82, No. 399. P. 918–924.
6. Лемешко Б.Ю., Лемешко С.Б. О сходимости распределений статистик и мощности критериев однородности Смирнова и Лемана–Розенблатта // Измерительная техника. 2005. № 12. С. 9–14.
7. Лемешко Б.Ю., Лемешко С.Б., Постовалов С.Н., Чимитова Е.В. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход. Новосибирск : Изд-во НГТУ, 2011.
8. Лемешко Б.Ю. Непараметрические критерии согласия. Руководство по применению. М. : ИНФРА-М, 2014. DOI: 10.12737/11873.
9. Лемешко Б.Ю. Критерии проверки отклонения распределения от нормального закона. Руководство по применению. М. : ИНФРА-М, 2015. DOI: 10.12737/6086.
10. Лемешко Б.Ю., Блинов П.Ю. Критерии проверки отклонения распределения от равномерного закона. Руководство по применению. М. : ИНФРА-М, 2015. DOI: 10.12737/11304.
11. Смирнов Н.В. Оценка расхождения между эмпирическими кривыми распределения в двух независимых выборках // Бюл. МГУ. Сер. А. 1939. Т. 2, № 2. С. 3–14.
12. Макаров А.А., Симонова Г.И. Исследование мощности двухвыборочного критерия Андерсена–Дарлинга в случае засорения одной из выборок // Статистические методы оценивания и проверки гипотез : межвуз. сб. науч. тр. № 20. Перм. ун-т. Пермь, 2007. С. 40–52.
13. Постовалов С.Н. Применение компьютерного моделирования для расширения прикладных возможностей классических методов проверки статистических гипотез : дис. ... д-ра техн. наук. Новосибирск, 2013.
14. Kiefer J. K-Sample Analogues of the Kolmogorov-Smirnov and Cramer-v. Mises Tests // Annals of Mathematical Statistics. 1959. V. 30, No. 2. P. 420–447.
15. Лемешко Б.Ю. Критерии проверки гипотез об однородности. Руководство по применению. М. : ИНФРА-М, 2017.
16. Zhang J. Powerful goodness-of-fit and multi-sample tests : PhD Thesis. Toronto : York University, 2001. URL: http://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/ftp05/NQ66371.pdf (дата обращения: 26.01.2016).
17. Zhang J. Powerful Two-Sample Tests Based on the Likelihood Ratio // Technometrics. 2006. V. 48, No. 1. P. 95–103.

Лемешко Борис Юрьевич, д-р техн. наук, профессор. E-mail: Lemeshko@ami.nstu.ru
Лемешко Станислав Борисович, канд. техн. наук. E-mail: skyer@mail.ru
Веретельникова Ирина Викторовна. E-mail: ira-veterok@mail.ru
Новосибирский государственный технический университет
Поступила в редакцию 10 июня 2017 г.

Lemeshko Boris Yu., Lemeshko Stanislav B., Veretel`nikova Irina V. (Novosibirsk State Technical University, Russian Federation).
Application of criteria for testing homogeneity of distribution laws.
Keywords: homogeneity tests; Smirnov test; Lehmann–Rosenblatt test; Anderson–Darling test; power of test.

DOI: 10.17223/19988605/41/3
The necessity of general homogeneity hypothesis checking, i.e. whether two (or more) samples of random variables belong to the same general aggregate, arises constantly in the course of analysis of unbiased errors of measuring instruments. Such task arises naturally in the course of measuring instruments calibration and when comparing lab tests results. Technologists, medical researchers and biologists also face the same task when processing results of experimental research.

The present paper deals with the Smirnov, the Lehmann-Rosenblatt and the Anderson-Darling two-samples homogeneity tests; statistics for these tests is given; advantages and disadvantages of the tests are discussed.

As a drawback of the Smirnov test we may mention a substantial discreteness of statistics distribution which must be taken into account when having significant values of n and equal number of m and n samples. This drawback can be overcome if we choose mu-

tually prime integers as m and n. But even when this drawback has been remedied the real statistics distribution still differs substantially from the Kolmogorov limit distribution as the checked hypothesis is true. Hence when using the latter for evaluation of the significance level (p_{value}) incorrect conclusions can be reached. One can avoid this only by modifying the test statistics.

As opposed to the Smirnov test, the Lehmann-Rosenblatt and the Anderson-Darling statistics distribution homogeneity tests do not actually differ from their limit distributions when having samples values of $m, n \geq 25$.

Comparative analysis of tests powers under discussion, conducted with the help of statistic modeling methods, showed that, as a rule, the Anderson-Darling test boasts of bigger power than the Lehmann-Rosenblatt test, especially in case of differences of samples in their measure of dispersion. At the same time, when having rather similar yet competing hypothesis and smaller number of samples, the Lehmann-Rosenblatt test can show advantage in power. The Smirnov test yields to the Lehmann-Rosenblatt and the Anderson-Darling competing tests, but in certain cases it can be quite competitive.

Previously, information on distribution of statistics for *k*-sample Anderson-Darling homogeneity test has been available only within a limited table of critical limits. In the present case the study of statistics distribution through statistic modeling methods with actual k values showed the presence of corresponding limit distributions. Results of such modeling showed that when using this test we can disregard the difference between the statistics distributions and the corresponding limit values while the number of samples $n_i \geq 30$.

Based on the results of statistics modeling, approximate models for limit distributions of the Anderson-Darling *k*-samples test for $k = 2 \div 11$. This paper shows the models created, represented by the laws of beta-distributions of the IIIrd with particular parameter values. Such models produced with the help of the Anderson-Darling *k*-sample test enable finding the values of p_{value}, thus making the results of statistics conclusions more informative and more substantiated.

The possibility of using the *Z* hang *k*-sample homogeneity test with statistics of Z_K, Z_C and Z_A is discussed, these being the extension of the Smirnov, the Lehmann-Rosenblatt and the Anderson-Darling test, respectively. The Zhang tests have some advantages in power with reference to the scale alternatives, but they somewhat yield to the Smirnov, the Lehmann-Rosenblatt and the Anderson-Darling tests with reference to the shift alternatives.

REFERENCES

1. Bolshev, L.N. & Smirnov, N. V. (1983) *Tablitsy matematicheskoy statistiki* [Tables for Mathematical Statistics]. Moscow: Nauka.
2. Lehmann, E.L. (1951) Consistency and unbiasedness of certain nonparametric tests. *The Annals of Mathematical Statistics*. 22(1). pp. 165–179.
3. Rosenblatt, M. (1952) Limit theorems associated with variants of the von Mises statistic. *The Annals of Mathematical Statistics*. 23. pp. 617–623.
4. Pettitt, A.N. (1976) A two-sample Anderson-Darling rank statistic. *Biometrika*. 63(1). pp. 161–168. DOI: 10.1093/biomet/63.1.161
5. Scholz, F.W. & Stephens M.A. (1987) K-Sample Anderson–Darling Tests. *Journal of the American Statistical Association*. 82(399). pp. 918–924. DOI: 10.1080/01621459.1987.10478517
6. Lemeshko, B.Yu. & Lemeshko, S.B. (2005) Statistical distribution convergence and homogeneity test power for Smirnov and Lehmann–Rosenblatt tests. *Measurement Techniques*. 48(12). pp. 1159–1166. DOI: 10.1007/s11018-006-0038-3
7. Lemeshko, B.Yu., Lemeshko, S.B., Postovalov, S.N. & Chimitova, E.V. (2011) *Statisticheskiy analiz dannykh, modelirovanie i issledovanie veroyatnostnykh zakonomernostey. Komp'yuternyy podkhod* [Statistical Data Analysis, Simulation and Study of Probability Regularities. Computer Approach]. Novosibirsk: NSTU.
8. Lemeshko, B.Yu. (2014) *Neparametricheskie kriterii soglasiya. Rukovodstvo po primeneniyu* [Nonparametric goodness-of-fit tests]. Moscow: INFRA–M. DOI: 10.12737/11873
9. Lemeshko, B.Yu. (2015) *Kriterii proverki otkloneniya raspredeleniya ot normal'nogo zakona* [Tests for checking the deviation from normal distribution law]. Moscow: INFRA–M. DOI: 10.12737/6086
10. Lemeshko, B.Yu. & Blinov, P.Yu. (2015) *Kriterii proverki otkloneniya raspredeleniya ot ravnomernogo zakona* [Tests for checking the deviation from uniform distribution law]. Moscow: INFRA–M. DOI: 10.12737/11304
11. Smirnov, N.V. (1939) Otsenka raskhozhdeniya mezhdu empiricheskimi krivymi raspredeleniya v dvukh nezavisimykh vyborkakh [Evaluation of discrepancy between the empirical distribution curves in two independent samples]. *Byuleten. MGU. Ser. A. – Bulletin of MGU, Series A*. 2(2). pp. 3–14.
12. Makarov, A.A. & Simonova, G.I. (2007) Issledovanie moshchnosti dvukhvyborochnogo kriteriya Andersena–Darlinga v sluchae zasoreniya odnoy iz vyborok [Studies of the power of the two-sample Anderson-Darling test in the case of littering of one of the samples]. In: Lumelsky, Ya.P. (ed.) *Statisticheskie metody otsenivaniya i proverki gipotez* [Statistics methods of evaluation and checking of hypothesis]. Perm: Perm State University. pp. 40–52.
13. Postovalov, S.N. (2013) *Primenenie komp'yuternogo modelirovaniya dlya rasshireniya prikladnykh vozmozhnostey klassicheskikh metodov proverki statisticheskikh gipotez* [Using of computer modeling for expanding application of classical methods of statistics hypothesis checking]. Engineering Dr. Diss. Novosibirsk.
14. Kiefer, J. (1959) K-Sample Analogues of the Kolmogorov-Smirnov and Cramer-v. Mises Tests. *Annals of Mathematical Statistics*. 30(2). pp. 420–447.
15. Lemeshko, B.Yu. (2017) *Kriterii proverki gipotez ob odnorodnosti* [Tests for homogeneity]. Moscow: INFRA–M.
16. Zhang, J. (2001) *Powerful goodness-of-fit and multi-sample tests*. PhD Thesis. Toronto: York University. [Online] Available from: http://www.collectionscanada.gc.ca/obj/s4/f2/dsk3/ftp05/NQ66371.pdf. (Accessed: 26th January 2016).
17. Zhang, J. (2006) Powerful Two-Sample Tests Based on the Likelihood Ratio. *Technometrics*. 48(1). pp. 95–103. DOi: 10.1198/004017005000000328. 95