УДК 524.19 Вестник СПбГУ. Математика. Механика. Астрономия. 2023. Т. 10 (68). Вып. 2
MSC 62K05

Об асимптотической мощности одного метода проверки гипотез о равенстве распределений*

В. Б. Мелас

Санкт-Петербургский государственный университет,
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: *Мелас В. Б.* Об асимптотической мощности одного метода проверки гипотез о равенстве распределений // Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия. 2023. Т. 10 (68). Вып. 2. С. 249–258.
https://doi.org/10.21638/spbu01.2023.206

В статье исследована асимптотическая мощность метода проверки гипотез о равенстве двух распределений, который можно рассматривать как обобщение критерия Манна — Уитни — Вилкоксона. Рассматривается класс распределений таких, что математическое ожидание квадрата некоторой вспомогательной функции конечно. Для случая, когда альтернативное распределение отличается от нулевого только величиной параметра сдвига, в явном виде найдены асимптотическое распределение критерия и асимптотическая мощность критерия. Ранее мощность этого критерия исследовалась только на основе статистического моделирования.

Ключевые слова: проверка гипотез о равенстве двух распределений, асимптотическая мощность статистических тестов, нормальное распределение, распределение Коши.

1. Введение. В статье найдены асимптотическое распределение и асимптотическая мощность критерия проверки гипотез о равенстве двух распределений, предложенного в работах [1, 2], для случая, когда альтернативное распределение отличается от нулевого только величиной параметра сдвига.

2. Постановка проблемы. Рассмотрим классическую задачу проверки гипотез о равенстве двух распределений:

$$H_0 \,:\, F_1 = F_2 \tag{1}$$

против альтернативы

$$H_1 \,:\, F_1 \neq F_2 \tag{2}$$

в случае двух независимых выборок $X = (X_1, \ldots, X_n)$ и $Y = (Y_1, \ldots, Y_m)$ с функциями распределения F_1 и F_2 соответственно.

Хорошо известно (см. [3]), что в случае, когда два распределения отличаются только средними и являются нормальными, классический тест Стьюдента обладает несколькими оптимальными свойствами. Если распределения не являются нормальными, но все еще различаются только средними, то вместо теста Стьюдента часто используется широко известная U-статистика Уилкоксона — Манна — Уитни (WMW). Однако можно показать, что если две нормальные выборки отличаются только дисперсиями, то мощность теста WMW очень мала. Для произвольных

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (грант № 20-01-00096-а).

распределений существуют универсальные методы, такие как тесты Колмогорова—Смирнова и Крамера—фон Мизеса (см. [4]), и тест Андерсона—Дарлинга (см. [5]), которые могут быть применены, но во многих случаях эти тесты имеют низкую мощность [1, 2].

Рассмотрим следующий тест:

$$\Phi_{nm} = \Phi_{AB} - \Phi_A - \Phi_B, \tag{3}$$

где

$$\Phi_A = \frac{1}{n^2} \sum_{1 \le i < j \le n} g(X_i - X_j), \Phi_B = \frac{1}{m^2} \sum_{1 \le i < j \le m} g(Y_i - Y_j),$$

$$\Phi_{AB} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} g(X_i - Y_j),$$

$g(x)$ есть некоторая заданная функция вещественной переменной. Этот тест был, по-видимому, впервые введен в работе [1], но его мощность исследовалась только с помощью статистического моделирования и лишь для случая $g(x) = \ln(|x|)$.

В работе [2] рассматривался случай

$$g(x) = \ln\bigl(1 + x^2\bigr).$$

Эта функция является с точностью до знака и постоянного слагаемого логарифмом плотности стандартного распределения Коши. Было показано в [2] с помощью статистического моделирования, что критерий (3) с такой функцией $g(x)$ имеет примерно такую же мощность, как и критерии Вилкоксона, Андерсона—Дарлинга и Колмогорова—Смирнова при альтернативе, которая отличается только величиной параметра сдвига, но значительно превосходит эти критерии, если есть различие в параметре масштаба. Предположим, что функции распределения F_1 и F_2 таковы, что для соответствующих случайных величин ξ выполняется неравенство

$$E[g(\xi)^2] < \infty, \tag{4}$$

где $g(x)$ — некоторая заданная функция. Рассмотрим класс распределений, задаваемых свойством (4). Обратим внимание, что если параметры известны, то тест, основанный на отношении правдоподобия, является самым мощным среди тестов с заданными параметрами. Предложенный выше тест при $g(x) = \ln\bigl(1 + x^2\bigr)$ можно рассматривать как приближение логарифма этого соотношения для распределения Коши. Это позволяет ожидать высокую эффективность теста для функций распределения, удовлятворяющих свойству (4) с такой функцией $g(x)$. В настоящей работе мы изучаем асимптотическую мощность теста (3) в случае функций $g(x), x \in R$ общего вида для альтернативных распределений, отличающихся от нулевого только величиной параметра сдвига.

3. Аналитическое исследование асимптотической мощности. Рассмотрим случай двух распределений, обладающих свойством (4) и отличающихся только величиной параметра сдвига.

Пусть $f(x)$ обозначает плотность F_1,

$$J(h, n) = \int_R g(x - y - |h|/\sqrt{n}) f(x) f(y) dxdy,$$

$$J_1 = J(0,n), J_2 = \int_R g^2(x-y)f(x)f(y)dxdy,$$
$$J_3 = \int_R g(x-y)g(x-z)f(x)f(y)f(z)dxdydz,$$

где h — произвольное заданное число. Пусть $g(x)$ — функция, симметричная относительно некоторой точки. Будем предполагать, что эта функция неотрицательна, симметрична относительно начала координат и дважды непрерывно дифференцируема. Тогда существует предел $J^*(h) = \lim_{n\to\infty} n(J(h,n) - J(0,n))$. Обозначим $\bar{b} = \sqrt{J^*(h)/h^2}$. Для упрощения обозначений будем рассматривать случай $n = m$. Общий случай рассматривается аналогичным образом. Обозначим

$$T_n = \Phi_{nn}.$$

Основным результатом настоящей работы является следующая теорема, которая устанавливает вид предельного распределения величины nT_n и представление для асимптотической эффективности теста.

Теорема 3.1. *Рассмотрим задачу проверки гипотезы (1)–(2), где обе функции обладают свойством (4) и имеют плотности распределения, симметричные относительно некоторой точки. Тогда:*

(i) при условии $n \to \infty$ функция распределения nT_n сходится при H_0 к функции распределения случайной величины

$$(aL)^2 + c, \tag{5}$$

где L имеет стандартное нормальное распределение, $a^2 = \sqrt{J_2 + J_1^2 - 2J_3}$, $c = J_1 - a^2$;

(ii) пусть $F_1(x) = F(x), F_2 = F(x + h/\sqrt{n})$, где F — произвольная функция распределения, с плотностью $f(x)$, симметричной относительно некоторой точки, и обладающая свойством (3), h — произвольное заданное число. Тогда функция распределения nT_n сходится при H_1 к функции распределения случайной величины

$$(aL + b)^2 + c,$$

где $b = \bar{b}h$. В этом случае мощность критерия nT_n с уровнем значимости α асимптотически равна

$$Pr\{L \geq z_{1-\alpha/2} - \bar{b}h/a\} + Pr\{L \leq -z_{1-\alpha/2} - \bar{b}h/a\},$$

где $z_{1-\alpha/2}$ является таким, что

$$Pr\{L \geq z_{1-\alpha/2}\} = \alpha/2.$$

Прежде чем доказывать теорему, проиллюстрируем ее применение на примерах[1]. Рассмотрим случай $g(x) = \ln(1 + x^2)$, который уже изучался в работе [2]. Непосредственная проверка показывает, что условие (4) выполняется для нормального распределения и распределения Коши среди многих других.

[1] Примеры построены магистром математико-механического факультета СПбГУ Анной Белковой.

Пример 3.1. *Пусть $f(x)$ — функция плотности стандартного нормального распределения. Численное интегрирование дает следующие результаты:*

$$J_1 = 0.810113, J_2 = 1.155022, J_3 = 0.763368, \bar{b} = 0.477.$$

Вычисляя коэффициент a по формуле из теоремы 3.1, получаем $a = 0.7303767$. В табл. 1 представлены теоретические значения мощности и эмпирические мощности, полученные в результате численной обработки данных $N = 1000$ повторений статистического моделирования двух выборок размера $n = 100$. Критическое значение критерия nT_n вычислялось с помощью 700 случайных перестановок.

Таблица 1. **Мощности для нормального распределения**

h	1.0	2.0	3.0	4.0	5.0	6.0
Эмпирическая мощность	0.100	0.264	0.487	0.726	0.892	0.969
Теоретическая мощность	0.100	0.257	0.499	0.742	0.904	0.974

Заметим, что эмпирическая мощность равна числу повторений численного эксперимента, в которых альтернативная гипотеза, будучи верной, принимается, деленному на общее число повторений. При $N \to \infty$ распределение этой случайной величины стремится к нормальному распределению. Стандартная проверка показывает, что во всех рассмотренных случаях различие значений эмпирической и теоретической мощности статистически незначимо.

Пример 3.2. *Пусть $f(x)$ — плотность стандартного распределения Коши. В этом случае в работе [2] с помощью таблиц интегралов показано, что*

$$J_1 = \ln 9, \bar{b} = \frac{1}{3}.$$

Численным интегрированием находим

$$J_2 = 9.577512,\ J_3 = 6.881056.$$

По теореме 3.1 получаем $a = 0.8955417$. Теоретические и эмпирические значения мощности представлены в табл. 2. Эмипирические мощности вычислялись тем же способом, что и в примере 3.1.

Таблица 2. **Мощности для распределения Коши**

h	1. 0	2.0	4.0	6.0	7.0	9.0
Эмпирическая мощность	0.070	0.117	0.258	0.546	0.677	0.888
Теоретическая мощность	0.066	0.115	0.319	0.607	0.740	0.917

Заметим, что при $h = 1.0, 2.0, 4.0$ различие значений эмпирической и теоретической мощностей статистически незначимо, что проверяется стандартным способом, исходя из замечания, сделанного при анализе результатов в предыдущей таблице.

Доказательство теоремы. Сначала докажем, что теорема верна для случая $g(x) = x^2$. Рассмотрим тест (3) с функцией $g(x) = x^2$.

Лемма 3.1. *Если условие (4) выполняется для $g(x) = x^2$, то справедливо следующее тождество:*

$$\Phi_{nn} = (\bar{x} - \bar{y})^2,$$

где

$$\bar{x} = (\sum_{i=1}^{n} X_i)/n, \bar{y} = (\sum_{i=1}^{n} Y_i)/n.$$

При $n \to \infty$ в случае выполнения гипотезы H_0 распределение величины $\sqrt{n\Phi_{nn}}$ сходится к нормальному распределению с дисперсией J_1 и нулевым математическим ожиданием. Если выполнена гипотеза H_1, то распределение этой величины сходится к нормальному с дисперсией J_1 и математическим ожиданием $\sqrt{2}h$.

Доказательство этой леммы можно найти в работе [2]. Для удобства читателя и в связи с тем, что этот результат имеет ключевое значение для доказательства теоремы, мы приводим доказательство этой леммы в Приложении.

В силу леммы 3.1 для $g(x) = x^2$ при выполнении гипотезы H_0 величина nT_n имеет вид (6) при $a^2 = J_1, c = 0$. Таким образом, утверждение (i) теоремы в этом случае справедливо. Для доказательства части (ii) теоремы заметим, что в силу леммы 3.1 при выполнении гипотезы H_1 величина nT_n имеет асимптотически то же распределение, что величина $(aL+b)^2$, где $b = \sqrt{2}h$. Непосредственным вычислением можно проверить, что $\bar{b} = \sqrt{2}$. Утверждение о мощности вытекает из того, что $(\sqrt{nT_n} - b)$ имеет нормальное распределение с дисперсией J_1.

Рассмотрим далее случай функции $g(x)$ общего вида. Пусть выполнена гипотеза H_0. Имеет место следующая лемма.

Лемма 3.2. *В условиях части (i) теоремы распределение величины nT_n сходится при $n \to \infty$ к распределению величины*

$$a^2L^2 + c, \tag{6}$$

где L имеет стандартное нормальное распределение, a и c — некоторые числа.

Доказательство леммы 3.2 дано в Приложении.

Обозначим предельное распределение величины nT_n как $R(L, a, c) = a^2L^2 + c$. В силу равенства

$$\lim_{n\to\infty} EnT_n = ER(L, a, c) = a^2 + c$$

мы получаем, применяя к величине nT_n закон больших чисел для U-статистик (см. [6]),

$$c = J_1 - a^2.$$

Заметим, что величину $(nT_n)^2$ можно представить как

$$\frac{1}{n^2}\{\sum_{i,j=1,\dots,n} (g(X_i - Y_j) - J_1) - \sum_{1\le i<j\le n} (g(X_i - X_j) - J_1) - \sum_{1\le i<j\le n} (g(Y_i - Y_j) - J_1) + nJ_1\}^2. \tag{7}$$

Используя формулу (7), представим предел $(nT_n)^2$ при $n \to \infty$ в виде

$$\lim_{n\to\infty} (nT_n)^2 = J_1^2 + \lim_{n\to\infty} (\hat{J}_1(n) - J_1)^2 + \lim_{n\to\infty} \frac{1}{n^2}\{2n(\hat{J}_1(n) - J_1)V_n + V_n^2\}, \tag{8}$$

где

$$\hat{J}_1(n) = \sum_{i=1}^{n} \frac{1}{n} g(X_i - Y_i),$$

$$V_n = \{2 \sum_{1 \le i < j \le n} (g(X_i - Y_j) - J_1) - \sum_{1 \le i < j \le n} (g(X_i - X_j) - J_1) - \sum_{1 \le i < j \le n} (g(Y_i - Y_j) - J_1)\}.$$

Прямое вычисление с использованием закона больших чисел для U-статистик показывает, что $\lim_{n \to \infty} \frac{1}{n} V_n = E \frac{1}{n} V_n = 0$. Таким образом,

$$\lim_{n \to \infty} (nT_n)^2 = J_1^2 + \lim_{n \to \infty} \frac{1}{n^2} V_n^2.$$

Заметим, что V_n^2 является алгебраической суммой произведений вида

$$R_{(i,j),(s,k)} = (g(Z_i - Z_j) - J_1)(g(Z_s - Z_k) - J_1),$$

где $Z = (Z_1, \dots, Z_{2n}) = (X_1, \dots, X_n, Y_1, \dots, Y_n)$.

В силу независимости и одинаковой распределенности величин $X_i, Y_j, i, j = 1, \dots, n$ получаем в силу закона больших чисел, что если все индексы различны, то

$$ER_{(i,j),(s,k)} = 0, i, j, s, k \in \{1, \dots, 2n\}.$$

Если пары индексов совпадают, $(i, j) = (s, k)$, то получаем

$$ER_{(i,j),(i,j)} = J_2 - J_1^2.$$

В случае, когда совпадает только один из индексов, получаем с помощью предельной теоремы для U-статистик ([6]), что

$$ER_{(i,j),(i,k)} = J_3 - J_1^2, \tag{9}$$

если $j \ne i, k \ne i, j$. Такое же соотношение выполняется для аналогичных случаев. Соответствующие произведения будем называть произведениями третьего вида.

Заметим, что число произведений вида $R_{(i,j),(i,j)}$ равно $2n(n-1)$. А число сумм третьего вида, входящих в V_n^2 со знаком минус, равно $4n(n-1)^2$, а число произведений со знаком плюс равно $4n(n-1)(n-2)$. Таким образом, математическое ожидание алгебраической суммы таких произведений равно $-4n(n-1)(J_3 - J_1^2)$.

Следовательно,

$$E \lim_{n \to \infty} (nT_n)^2 = J_1^2 + E \lim_{n \to \infty} (\frac{1}{n^2} V_n) = J_1^2 + 2(J_2 - J_1^2) - 4(J_3 - J_1^2).$$

Из соотношения

$$E \lim_{n \to \infty} (nT_n)^2 = E(R(L, a, c))^2 = 2a^4 + J_1^2,$$

получаем, переходя к пределу при $n \to \infty$, что $a^2 = \sqrt{J_2 + J_1^2 - 2J_3}$.

Пусть теперь H_1 имеет место. Повторяя проведенные аргументы, мы можем проверить, что функция распределения величины nT_n сходится при выполнении гипотезы H_1 и при $n \to \infty$ к функции распределения величины

$$(aL + b)^2 + c,$$

где b есть некоторое число, a и c те же самые, что в части (i). Заметим, что при выполнении H_1 величина EnT_n асимптотически эквивалентна величине

$$(n(J_h - J_0))^2 + En\hat{T}_n,$$

где $\hat{T}_n$ получается из T_n заменой Y_i на $Y_i - b/\sqrt{n}$, $i = 1, \dots, n$, и мы получаем, переходя к пределу при $n \to \infty$, что

$$b = \bar{b}h, \bar{b} = \sqrt{J^*(h)/h^2}.$$

Асимптотическое поведение мощности, объявленное в (ii), вытекает из асимптотической нормальности $\sqrt{nT_n - c}$, что завершает доказательство теоремы.

4. Приложение. Доказательство леммы 3.1. Обозначим

$$Z = (X, Y) = (X_1, \dots, X_n, Y_1, \dots, Y_n), V(Z) = \frac{1}{2} \sum_{i=1}^{2n} \sum_{j=1}^{2n} (Z_i - Z_j)^2.$$

Доказательство следует из известной формулы (см., например, [6, p. 296])

$$\frac{1}{n(n-1)} \sum_{1 \le i < j \le n} (X_i - X_j)^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{x})^2 \tag{10}$$

и очевидного тождества

$$\sum_{i=1}^{2n} \sum_{j=1}^{2n} (Z_i - Z_j)^2 = \sum_{i,j=1}^{n} (X_i - X_j)^2 + \sum_{i,j=1}^{n} (Y_i - Y_j)^2 + 2 \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - Y_j)^2, \tag{11}$$

путем прямых, но нетривиальных вычислений.

Действительно, давайте использовать стандартную форму записи:

$$S_x^2 = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{x})^2,$$

S_y^2 и S_z^2 будем понимать аналогичным образом. Обозначим

$$S_{xy} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - Y_j)^2.$$

Используя формулы (10), получаем

$$V(Z) = 2n[\sum_{i=1}^{n} (X_i - (\bar{x} + \bar{y})/2)^2 + \sum_{j=1}^{n} (Y_i - (\bar{x} + \bar{y})/2)^2] = 2n(n-1)(S_x^2 + S_y^2) + n^2(\bar{x} - \bar{y})^2. \tag{12}$$

Из (10) и (11) получаем

$$n^2 S_{xy} = V(Z) - n(n-1)(S_x^2 + S_y^2). \tag{13}$$

Следовательно,

$$S_{xy} = \frac{1}{n}(n-1)(S_x^2 + S_y^2) + (\bar{x} - \bar{y})^2,$$

и мы получаем

$$\Phi_{nn} = S_{xy} - \frac{1}{n}(n-1)(S_x^2 + S_y^2) = (\bar{x} - \bar{y})^2.$$

Пусть справедлива гипотеза H_0. По классической центральной предельной теореме величина $\sqrt{n\Phi_{nn}}$ имеет распределение, сходящееся при $n \to \infty$ к нормальному распределению с нулевым математическим ожиданием и дисперсией J_1. Последнее утверждение леммы проверяется прямым вычислением. Таким образом, лемма 3.1 доказана. Из этой леммы следует, что критерий Φ_{nn} в этом случае эквивалентен критерию $(\bar{x} - \bar{y})^2$. $\square$

Доказательство леммы 3.2. Обозначим

$$U_n(u, g) = \left(\begin{array}{c} n \\ 2 \end{array} \right)^{-1} \sum_{1 \le u_i < u_j \le n} g(u_i - u_j), \; u = (u_1, \ldots, u_n). \tag{14}$$

Функция $U_n(u, g)$ есть по определению (см. [6]) U-статистика. Напомним, что мы приняли $m = n$ и

$$\Phi_{AB} = \Phi_{AB}(X, Y, g) = -\frac{1}{n^2} \sum_{i,j=1}^{n} g(X_i - Y_j),$$

$$\Phi_A(X, g) + \Phi_B(Y, g) = -\frac{1}{n^2} \sum_{1 \le i < j \le n} g(X_i - X_j) - \frac{1}{n^2} \sum_{1 \le i < j \le n} g(Y_i - Y_j).$$

Следовательно,

$$\Phi_A(X, g) = -\frac{1}{2}\frac{n-1}{n} U_n(X, g), \;\; -\Phi_B(Y, g) = \frac{1}{2}\frac{n-1}{n} U_n(Y, g), \tag{15}$$

$$\Phi_{AB}(X, Y, g) = \frac{1}{n^2}\left(\begin{array}{c} 2n \\ 2 \end{array} \right) U_{2n}(Z, g) - \frac{1}{2}\frac{n-1}{n} U_n(X, g) - \frac{1}{2}\frac{n-1}{n} U_n(Y, g), \tag{16}$$

где $Z = (Z_1, \ldots, Z_{2n}) = (X_1 \ldots, X_n, Y_1, \ldots, Y_n)$. Применим предельную теорему (см. теорему 7.1 [6]) к каждому из выражений $\Phi_A(X, g), \Phi_B(Y, g)$ и $\Phi_{AB}(X, Y, g)$. Прямое вычисление показывает, что условие невырожденности выполняется, если выполнено условие (4). Пусть сначала условие (4) выполняется для $g(x) = g^*(x) = x^2$. По предельной теореме [6] величины $n\Phi_A(X, g)$ и $n\Phi_A(X, g^*)$ имеют в пределе нормальное распределение. Так как нормальные распределения полностью определяются параметрами сдвига и масштаба, то имеет место равенство

$$\frac{1}{n} \sum_{1 \le i < j \le n} g(X_i - X_j) = a^2 \frac{1}{n} \sum_{1 \le i < j \le n} g^*(X_i - X_j) + \tilde{c} + \tilde{\eta}_n, \tag{17}$$

где a и $\tilde{c}$ — некоторые числа, а величина $\tilde{\eta}_n$ есть случайная величина, сходящаяся по распределению к постоянной, равной нулю. И так как величины $X_i - X_j$ и $Y_i - Y_j$, $1 \le i < j \le n$ имеют одно и то же распределение, получаем

$$\frac{1}{n} \sum_{1 \le i < j \le n} g(Y_i - Y_j) = a^2 \frac{1}{n} \sum_{1 \le i < j \le n} g^*(Y_i - Y_j) + \tilde{c} + \tilde{\eta}_n \tag{18}$$

с теми же самыми постоянными a и $\tilde{c}$, что и в равенстве (17).

По той же причине и принимая во внимание (15) для Φ_{AB}, мы получаем формулу

$$\frac{1}{n^2} \sum_{i,j=1}^{n} g(X_i - Y_j) = a \frac{1}{n^2} \sum_{i,j=1}^{n} g^*(X_i - Y_j) + \bar{\eta}_n + \bar{c},$$

где постоянная a такая же, как в (17), но $\bar{c} \neq \tilde{c}$ и $\bar{\eta}_n \neq \tilde{\eta}_n$. И мы получаем

$$nT_n(X,Y,g) = a^2 nT_n(X,Y,g^*) + c + \eta_n,$$

где η_n сходится по вероятности к 0, a и c есть постоянные, a есть то же самое, что в формуле (17), а $g^*(x) = x^2$. По лемме 3.1 $\frac{1}{J_1} nT_n(X,Y,g^*)$ сходится по распределению к L^2. Таким образом, предельное распределение $nT_n(X,Y,g)$ имеет вид $a^2L^2 + c$.

Рассмотрим теперь случай, когда условие (4) не выполняется для $g(x) = g^*(x)$.

Пусть K — произвольное положительное число,

$$\tilde{X} = (\tilde{X_1}, \ldots, \tilde{X_n}), \;\; \tilde{Y} = (\tilde{Y_1}, \ldots, \tilde{Y_n}),$$

где $\tilde{X}_i = X_i$, если $|X_i| \leq K$ и $\tilde{X}_i = K$ и $X_i > 0$, $\tilde{X}_i = -K$ $X_i < 0$. Пусть $\tilde{Y}_i$ определены подобным образом. Теперь условие (4) выполняется и для заданной функции $g(x)$, и для $g(x) = x^2$ (в силу конечности дисперсии модифицированных величин).

Рассмотрим величину

$$n\{\frac{1}{n^2}\sum_{i,j=1}^{n} g(\tilde{X}_i - \tilde{Y}_j) - \frac{1}{n^2}\sum_{i<j} g(\tilde{X}_i - \tilde{X}_j) - \frac{1}{n^2}\sum_{i<j} g(\tilde{Y}_i - \tilde{Y}_j)\}.$$

В силу представленных выше аргументов предельное распределение этой величины имеет вид $R(L,a,c)$. При $K \to \infty$ предельное распределение существует (по теореме 7.1 [6]) и имеет такой же вид. Таким образом, лемма 3.2 доказана. $\square$

5. Заключение. В данной работе получено асимптотическое распределение рассматриваемого критерия и найдена формула для асимптотической мощности. С помощью стастистического моделирования установлено, что найденная формула позволяет получать теоретические значения мощности, которые статистически незначимо отличаются от эмпирических мощностей, найденных с помощью моделирования. Полученные результаты могут быть использованы для определения рационального размера выборки, т. е. для планирования эксперимента по проверке гипотез. Найденные формулы будут также полезны для дальнейшего исследования рассматриваемого критерия. Например, для оптимального выбора вспомогательной функции.

Литература

1. Zech G., Aslan B. New test for the multivariate two-sample problem based on the concept of minimum energy. *Journal of Statistical Computation and Simulation* **75** (2), 109–119 (2005).

2. Melas V., Salnikov D. On Asymptotic Power of the New Test for Equality of Two Distributions. In: *Recent Developments in Stochastic Methods and Applications*, vol. 371, 204–214. Springer Proceedings in Mathematics and Statistics (2021).

3. Леман Э. Л. Проверка статистических гипотез, пер. с англ. Москва, Наука (1979).

4. Buening H. Kolmogorov — Smirnov and Cramer — von Mises type two-sample tests with various weight functions. *Communications in Statistics-Simulation and Computation* **30**, 847–865 (2001).

5. Anderson T. W., Darling D. A. A test of goodness-of-fit. *J. Am. Stat. Assoc.* **49**, 765–769 (1954).

6. Hoeffding W. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics* **19**, 293–325 (1948).

Статья поступила в редакцию 12 марта 2022 г.;
доработана 12 марта 2022 г.;
рекомендована к печати 17 ноября 2022 г.

Контактная информация:

Мелас Вячеслав Борисович — д-р физ.-мат. наук, проф.; vbmelas@yandex.ru

On the asymptotic power of a method for testing hypothesis about equality of distributions*

V. B. Melas

St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation

For citation: Melas V. B. On the asymptotic power of a method for testing hypothesis about equality of distributions. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy*, 2023, vol. 10 (68), issue 2, pp. 249–258. https://doi.org/10.21638/spbu01.2023.206 (In Russian)

The paper is devoted to studying the asymptotic power of a method for testing hypothesis on equality of two distributions that can be considered as a generalization of Mann — Whitney — Wilcoxon test. We consider a class of distributions such that the expectation of the square of an auxiliary function is finite. For the case when alternative distribution differ from the initial one only by a shift the asymptotic distribution and asymptotic power of the test are found explicitly. Up to now the power of the test was studied only by stochastic simulation.

Keywords: testing hypothesis on equality of two distributions, asymptotic power of statistical tests, Normal distribution, Cashy distribution.

References

1. Zech G., Aslan B. New test for the multivariate two-sample problem based on the concept of minimum energy. *Journal of Statistical Computation and Simulation* **75** (2), 109–119 (2005).

2. Melas V., Salnikov D. On Asymptotic Power of the New Test for Equality of Two Distributions. In: *Recent Developments in Stochastic Methods and Applications*, vol. 371, 204–214. Springer Proceedings in Mathematics and Statistics (2021).

3. Lehmann E. L. *Testing statistical hypothess.* New York, Yohn Wiley & Sons (1959) [Rus. ed.: Lehmann E. L. *Proverka statisticheskikh gipotez.* Moscow, Nauka Publ. (1979)].

4. Buening H. Kolmogorov — Smirnov and Cramer — von Mises type two-sample tests with various weight functions. *Communications in Statistics-Simulation and Computation* **30**, 847–865 (2001).

5. Anderson T. W., Darling D. A. A test of goodness-of-fit. *J. Am. Stat. Assoc.* **49**, 765–769 (1954).

6. Hoeffding W. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics* **19**, 293–325 (1948).

Received: March 12, 2022
Revised: March 12, 2022
Accepted: November 17, 2022

Author's information:

Vyacheslav B. Melas — vbmelas@yandex.ru

*This work was supported by the Russian Foundation for Basic Research (project no. 20-01-00096-a).