Об асимптотической мощности "энергетического" теста для проверки гипотез о равенстве двух распределений

В.Б. Мелас, Д.И. Сальников[2]

1. Введение

Найдены асимптотическое распределение и асимптотическая мощность критерия проверки гипотез о равенстве двух распределений, предложенного в работах [1], [2], для случая, когда альтернативное распределение отличается сдвигом и(или) параметром масштаба. Эти результаты являются обобщением и развитием результатов статьи [3]. Проведено сравнение асимптотической мощности с эмпирической мощностью, найденной с помощью метода Монте-Карло для слчаев, когда сравниваются между собой два нормальные распределения или два распределения Коши для различных размеров выборок.Численные эксперименты подтверждают близость эмпирической мощности к асимптотической.

2. Постановка проблемы

Рассмотрим 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.

Предположим, что функции распределения F_1 и F_2 принадлежат классу функций распределений случайных величин ξ, таких, что

$$E[g(\xi)^2] < \infty, \tag{3}$$

где g - некоторая заданная функция. Многие распределения, в том числе нормальное распределение и распределение Коши, обладают этим свойством при $g(x) = \ln(1 + x^2)$.

В случае, когда два распределения отличаются только сдвигом, во многих случаях наиболее мощным является тест Манна-Уитни- Вилкоксона. Однако хорошо известно, что этот тест не позволяет дискриминировать распределения, разлиичающиеся параметром масштаба (см. [1], [2]). Мы хотели бы иметь тест, подходящий для ситуаций, когда нулевое распределение относится к классу распределений, обладающих свойством (3) для функции g общего вида, а альтернативное распределение отличается только сдвигом и/или преобразованем масштаба. Задачи, в которых важно учитывать возможность различия в параметре масштаба возникают во многих практических областях применения, влючая физиологию и психологию (см. например недавнюю работу [4]). Рассмотрим следующий тест

$$\Phi_{nm} = \Phi_{nm}(X, Y) = \Phi_{AB} - \Phi_A - \Phi_B, \tag{4}$$

[2]St.-Petersburg State University, Russia, e-mail: vbmelas@yandex.ru

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

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

где $g(x)$ есть некоторая заданная функция. Будем предполагать, что эта функция неотрицательна, симметрична относительно начала координат и дважды непрерывно дифференцируема. Этот тест был, по-видимому, впервые введен в работе [1], и назван "энергетическим"тестом, но его мощность исследовалась только с помощью статистического моделирования и лишь для случая $g(x) = \ln(|x|)$.

В работе [2] критерий (4) изучался для случая

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

Было показано с помощью статистического моделирования, что критерий (4) с такой функцией g имеет для многих распределений примерно такую же мощность как лучший из критериев Вилкоксона, Андесона- Дарлинга и Колмогорова-Смирнова при альтернативе, которая отличается только величиной параметра сдвига, но значительно превосходит эти критерии, если есть различие в параметре масштаба. .

Рассмотрим класс распределений, задаваемых свойством (3). Асимптотическая мощность теста для функций распределения, удовлетворяющих свойству (3), была изучена в работе [3] в случае функций g общего вида для распределений, отличающихся только сдвигом.

В настоящей работе мы изучаем асимптотическую мощность теста (4) для альтернативных распределений, отличающися от нулевого величиной параметра сдвига и/или параметра масштаба.

3. Асимптотическая мощность

Рассмотрим случай двух распределений, обладающих свойством (3) и отличающихся сдвигом и(или) параметром масштаба.

Пусть $f(x)$ обозначает плотность F_1,

$$J(h_1, n) = \int_R g(x - y - h_1/\sqrt{n}) f(x) f(y) dx dy,$$

$$J_1 = J(0, n), J_2 = \int_R g^2(x - y) f(x) f(y) dx dy,$$

$$J_3 = \int_R g(x - y) g(x - z) f(x) f(y) f(z) dx dy dz$$

$$J1(h_2, n) = \int_R g(x - y(1 + h_2/\sqrt{n})) f(x) f(y) dx dy,$$

Заметим, что

$$\int_R g'(x - y) f(x) f(y) dx dy = 0, \tag{5}$$

так как функция g(x), по предположению, дифференцируема и симметрична относительно нуля. Обозначим

$$J^*(h_1) = h_1^2 \int_R g''(x-y) f(x) f(y) dx dy,$$

$$J1^*(h_2) = (1/2) \int_R g''_\beta(x - y(1+\beta h_2))|_{\beta=0} f(x) f(y) dx dy$$

Обозначим

$$b_1^2 = |J^*(h_1)|, \tag{6}$$

$$b_2^2 = |J1^*(h_2)| \tag{7}$$

Для упрощения обозначений будем рассматривать случай $n = m$. Общий случай рассматривается аналогичным образом. Обозначим

$$T_n = T_n(X,Y) = \Phi_{nn}(X,Y).$$

Основным результатом настоящей работы является следующая теорема, которая устанавливает вид предельного распределения величины nT_n и представление для асимптотической эффективности теста. Эта теорема является обобщением Теоремы 3.1 из работы ([3]).

Теорема 3.1 *Рассмотрим задачу проверки гипотезы (1)-(2), где обе функции обладают свойством (3) 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

$$(aL)^2, \tag{8}$$

где L - случайная величина, которая имеет стандартное нормальное распределение,

$$a^2 = \sqrt{J_2 + J_1^2 - 2J_3}, \tag{9}$$

(ii) Пусть $F_1(x) = F(x), F_2(x) = F(x(1 + h_2/\sqrt{n}) + h_1/\sqrt{n})$, где F — произвольная функция распределения,с плотностью $f(x)$ симметричной относительно некоторой точки и обладающая свойством (3), h_1, h_2 - произвольные заданные числа .
Тогда функция распределения nT_n сходится при выполнении гипотезы H_1 к распределению случайной величины

$$(aL + \sqrt{b_1^2 + b_2^2})^2, \tag{10}$$

где b_1 имеет вид (6), b_2 определено в (7), а задано формулой (9).

Мощность критерия nT_n с уровнем значимости α асимптотически равна

$$Pr\{L \geq z_{1-\alpha/2} - b/a\} + Pr\{L \leq -z_{1-\alpha/2} - b/a\}, \tag{11}$$

где $b = \sqrt{b_1^2 + b_2^2}$, $z_{1-\alpha/2}$ является таким, что

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

Прежде, чем доказывать теорему, проиллюстрируем её применение на двух примерах для случая, когда два распределения различаются только параметром масштаба. Рассмотрим случай $g(x) = \ln(1+x^2)$, который уже изучался в работе [2]. Непосредственная проверка показывает, что условие (3) выполняется для нормального распределения и распределения Коши среди многих других. На двух примерах мы демонстрируем, что асимптотические формулы дают хорошее соответствие эмпирическим оценкам мощности для случая, когда распределения различаются (только) параметром масштаба. Для этих же примеров, но с распределениями, различающимися только сдвигом, подобное соответсвие было установлено в статье [3].

Пример 3.1 *Нормальное распределение. Пусть $f(x)$ - функция плотности стандартного нормального распределения, $h_1 = 0$, $\alpha = 0.05$. Численное интегрирование дает следующие результаты*

$$J_1 = 0.810113, J_2 = 1.155022, J_3 = 0.763368.$$

Вычисляя коэффициент a по формуле из теоремы 3.1, получаем a = 0.7303767. Также b_2 вычисляем по формуле (7). В таблице 3.1 представлены теоретические значения мощности, вычисленные по формуле (11), и эмпирические мощности, полученные в результате численной обработки данных N= 1000 повторений статистического моделирования двух выборок размера n=100. Критическое значение критерия T_n вычислялось с помощью 700 случайных перестановок.

Таблица 1. Значение эмпирической (Э) и асимптотической (А) мощности для нормального распределения

h_2	1.0	2.0	3.0	4.0	5.0	6.0	7.0	9.0
Э.Мощность								
А.Мощность								

Пример 3.2 *Распределение Коши. Пусть $f(x)$ - плотность стандартного распределения Коши, $h_1 = 0$, $\alpha = 0.05$, n=100. 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, что a=0.8955417. Теоретические и эмпирические значения мощности представлены в таблице 3.2. Эмипирические мощности вычислялись тем же способом, что в Примере 3.1.

Доказательство теоремы. Сначала докажем, что теорема верна для случая $g(x) = \frac{1}{2}x^2$. В этом случае критерий может быть записан в более простом виде.

Лемма 3.1 *Рассмотрим тест (4) с функцией $g(x) = g^*(x) = \frac{1}{2}x^2$. Если условие (3) выполняется для $g(x) = g^*(x)$, то справедливо следующее тождество*

$$\Phi_{nn} = \frac{1}{2}(\bar{x} - \bar{y})^2$$

Таблица 2. Значение эмпирической (Э) и асимптотической (А) мощности для распределения Коши

h_2	1	2	3	4	5	6	7	8	9	10
(Э) n=100	6.6	10.5	17.9	26.9	37.3	49.5	59.8	68.1	77.2	83.8
(Э) n=400	4.3	10	19.3	29.7	41.4	54.4	67.9	78.5	86.1	91.5
(Э) n=900	6.2	10.9	21.6	32.1	45.4	61.6	75.3	85.8	91.9	95.5
(Э) n=1600	6.5	11	20.3	32.2	46.8	62.7	77.3	85.2	93.1	96.7
(А)	6.6	11.6	20.1	31.9	46.2	60.8	74	84.6	91.8	96.1

где

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

При $n \to \infty$ распределение величины $\sqrt{n\Phi_{nn}}$ сходится к нормальному распределению с дисперсией J_1 и математическим ожиданием 0 при выполнении гипотезы H_0.

Доказательство этой леммы можно найти в работе [3]. Для удобства читателя и в связи с тем, что этот результат имеет ключевое значение для доказательства теоремы, мы приводим полное доказательство этой леммы (которое отличается от доказательства в [3] небольшими деталями)в Приложении.

В силу Леммы 3.1 при $g(x) = \frac{1}{2}x^2$ при выполнении гипотезы H_0 величина nT_n имеет вид (12) при $a^2 = J_1$. Таким образом, утверждение (i) теоремы в этом случае справедливо. Для доказательства части (ii) теоремы заметим, что в силу Леммы 3.1 при $g(x) = \frac{1}{2}x^2$ при выполнении гипотезы H_1 величина nT_n имеет асимптотически то же распределение, что величина $(aL+b)^2$, где $b = h_1$. Непосредственным вычислением можно проверить, что $b_2 = 0$ для любого h_2. Утверждение о мощности вытекает из того, что $(\sqrt{nT_n} - b)$ имеет нормальное распределение с дисперсией $\frac{1}{2}J_1$.

Рассмотрим далее случай функции $g(x)$ общего вида. Имеет место

Лемма 3.2 *(i)В условиях части (I) теоремы 3.1 распределение величины nT_n сходится при $n \to \infty$ к распределению величины*

$$a^2 L^2, \tag{12}$$

где L имеет стандартное нормальное распределение, a - некоторырое число.

(ii) В условиях части (II) теоремы 3.1. распределение величины nT_n сходится при $n \to \infty$ к распределению величины

$$(aL+b)^2, b = b_1 + b_2, \tag{13}$$

где L имеет стандартное нормальное распределение, a то же самое, что в части (i),b_1 определено равенством (6) b_2 определено в (7).

Доказательство Леммы 3.2 дано в приложении.

Обозначим

$$R(L,a) = a^2 L^2,$$

предельное распределение величины nT_n. В силу равенства

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

мы получаем, применяя к величине nT_n закон больших чисел для U-статистик (см. [5]) Заметим, что величину $(nT_n)^2$ можно представить как

$$\frac{1}{n^2}\{\sum_{i,j=1,\dots,n}(g(X_i-Y_j)-J_1)-\sum_{1\le<i,j\le n}(g(X_i-X_j)-J_1)-\sum_{1\le<i,j\le n}(g(Y_i-Y_j)-J_1)+nJ_1\}^2. \tag{14}$$

Используя формулу (14), представим предел $(nT_n)^2$ при $n\to\infty$ в виде

$$\lim_{n\to\infty}(nT_n)^2 = J_1^2 + \lim_{n\to\infty}(\hat{J}_1(n)-J_1)^2 + \lim_{n\to\infty}\frac{1}{n^2}\{2n(\hat{J}_1(n)-J_1)V_n + V_n^2\}, \tag{15}$$

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 имеет место равенство

$$ER_{(i,j),(i,k)} = J_3 - J_1^2, \tag{16}$$

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 $4n(n-1)^2$, и число произведений со знаком плюс равно $4n(n-1)(n-2)$. Таким образом, математическое ожидание алгебраической суммы таких произведений равно $-4n(n-1)(J_3-J_1^2)$.

Таким образом,

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

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, тогда по Лемме 3.2, часть(ii) функция распределения величины nT_n сходится при выполнении H_1 при $n \to \infty$ к функции распределения величины

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

где $b = \sqrt{b_1^2 + b_2^2}$, a и c те же самые, что в части (i). Из этого представления вытекает формула для асимптотической мощности, 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, [5], p.296)

$$\frac{1}{n(n-1)}\sum_{1\le i<j\le n}(X_i - X_j)^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{x})^2 \tag{17}$$

и очевидного тождества

$$\sum_{i=1}^{2n}\sum_{j=1}^{2n}(Z_i - Z_j)^2 = \sum_{i,j=1}^{n}(X_i - X_j)^2 + \sum_{i,j=1}^{n}(Y_i - Y_j)^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - Y_j)^2, \tag{18}$$

путем прямых, но нетривиальных вычислений.

Действительно, давайте использовать стандартную форму записи

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

И 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 (17), получаем

$$V(Z) = 2n[\sum_{i=1}^{n}(X_i - (\bar{x}+\bar{y})/2)^2 + \sum_{j=1}^{n}(Y_i - (\bar{x}+\bar{y})/2)^2] = 2n(n-1)(S_x^2 + S_y^2) + n^2(\bar{x}-\bar{y})^2. \tag{19}$$

Из (17) и (18) получаем

$$n^2 S_{xy} = V(Z) - n(n-1)(S_x^2 + S_y^2). \tag{20}$$

Следовательно

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

и мы получаем

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

По классической центральной предельной теореме величина $\sqrt{n\Phi_{nn}}$ имеет распределение, сходящееся при выпонении H_0 для $n \to \infty$ к нормальному распределению с дисперсией J_1 и математическим ожиданием 0. Таким образом, лемма 3.1 доказана. Из этой леммы следует, что критерий Φ_{nn} в этом случае эквивалентен критерию $(\bar{x} - \bar{y})^2$.

Доказательство Леммы 3.2. Начнем с доказательства части (i). В этой части предполагается выполненной гипотеза H_0. Мы проведём доказательство в форме, которая допусквет естественное обобщение на случай, когда выполнена гипотеза H_1. Заметим, что при $g(x) = g^*(x) = x^2/2$ в силу Леммы 3.1 и центральной предельной теоремы при $n \to \infty$ распределение nT_n сходится к распределению величины

$$(aL + b)^2, \tag{21}$$

где $a = 1, b = h_2$ при выполнении гипотезы H_0 или гипотезы H_1, причем для гипотезы H_0 $b = 0$. Покажем, что этот факт остается верным для (симметричной дважды непрерывно дифференцируемой функции g(z)= g (x-y), $x, y \in R^1$ общего вида (при дополнительны условии, что эта функция является невырожденным ядром) для некоторых чисел a и b. Обозначим

$$U_n(u, g) = \left(\begin{array}{c} n \\ 2 \end{array} \right)^{-1} \sum_{1 \le u_i < u_j \le n} g(u_i - u_j), \ u = (u_1, \ldots, u_n). \tag{22}$$

Функция $U_n(u, g)$ есть по определению (см.[5]) 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,

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

$$\Phi_{AB}(X, Y, g) = \frac{1}{n^2}\left(\begin{array}{c} 2n \\ 2 \end{array} \right) U_{2n}(Z, g) - \frac{1}{2}\frac{n-1}{n} U_n(X, g) - \frac{1}{2}\frac{n-1}{n} U_n(Y, g), \tag{23}$$

where $Z = (Z_1, \ldots, Z_{2n}) = ((X_1 \ldots, X_n, Y_1, \ldots, Y_n)$. Применим предельную теорему (см. Теорему 7.1 [5]) 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. По этой теореме величины $n\Phi_A(X,g)$ and $n\Phi_A(X,g^*)$ имееют в пределе нормальное распределение. Так как нормальные распределения полностью определяются параметрами сдвига и масштаба, то имеет место равенство

$$\frac{1}{n}\sum_{1\le i<j\le n} g(X_i - X_j) = a^2\frac{1}{n}\sum_{1\le i<j\le n} g^*(X_i - X_j) + c + \tilde{\eta}_n, \tag{24}$$

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, это равенство имеет место при замене X на Y с теми же самыми постоянными a и c.
По той же причине и принимая во внимание (23), мы получаем формулу

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

где постояннst a и c такие же как в (24) и $\bar{c} = \tilde{c}$, но $\bar{\eta}_n \neq \tilde{\eta}_n$ 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, а то же самое, что в формуле (24). По Лемме 3.1 $\frac{1}{2}nT_n(X,Y,g^*)$ сходится по распределению к L^2. Таким образом предельное распределение $nT_n(X,Y,g)$ имеет вид a^2L^2.

Рассмотрим теперь случай, когда условие (3) не выполняется для $g(x) = \frac{1}{2}x^2$

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 $|X_i| \le K$ и $\tilde{X}_i = K$ и $X_i > 0$, $\tilde{X}_i = -K$ $X_i < 0$в противном случае. Пусть $\tilde{Y}_i$ определены подобным образом. Через $\tilde{f}$ обозначим плотность распределения $\tilde{X}_1$(которая есть также плотность распределения $\tilde{X}_i, i = 2,\ldots,n$ and $\tilde{Y}_i, i = 1,\ldots,n$). Теперь условие (3) очевидно выполняется для заданной функции g и для $g = g^*$ с функцией f(x) замененной на $\tilde{f}(x)$.

Рассмотрим величину

$$n\{\frac{1}{n^2}\sum_{i,j=1}^{n} g(\tilde{X}_i - \tilde{Y}_j) - \frac{1}{n^2}\sum_{i<j} g(\tilde{X}_i - \tilde{X}_j) - \tag{25}$$

$$\frac{1}{n^2}\sum_{i<j} g(\tilde{Y}_i - \tilde{Y}_j)\}. \tag{26}$$

В силу представленных выше аргументов предельное распределение этой величины имеет вид $R(L,a)$. При $K \to \infty$ предельное распределение существует (по теореме 7.1[5]) и имеет такой же вид. Таким образом, часть (i) Леммы 3.2 доказана.

Часть (ii) доказывается следующим образом. Обозначим

$$\tilde{Y} = (Y_1(1 + h_2/\sqrt{n}) + h_1/\sqrt{n}, \ldots, Y_n(1 + h_2/\sqrt{n}) + h_1/\sqrt{n}).$$

При выполнении гипотезы H_1 величины $\tilde{Y}_i, i = 1, 2, \dots, n$ - независимые случайные величины с функцией распределения $F_1(x)$. Заметим, что $nT_n(X, Y)$ имеет вид

$$nT_n(X,Y) = \frac{1}{n}\sum_{i=1}^{n} g((X_i - \tilde{Y}_i) - Y_i h_2/\sqrt{n} - h_1/\sqrt{n}) + \eta_n, \tag{27}$$

где η_n -случайная величина, такая что $E\eta_n \to 0$ при $n \to \infty/$ что вытекает из закона больших чисел для U-статистик.

Положим $\beta = \sqrt{n}$. Так как функция $g(x)$, по предположению, симметрична и дважды непрерывно дифференцируема, то

$$g((x-y) - yh_2/\sqrt{n} - h_1/\sqrt{n}) = g(x-y) + (yh_2 - h_1)\beta g'_\beta((x-y) + yh_2\beta - h_1\beta)|_{\beta=0}$$
$$+\frac{1}{2}((yh_2 - h_1)\beta)^2 g''_\beta((x-y) - h_2\beta - h_1\beta)|_{\beta=0} + o(\beta^2).$$

Вычисляя математическое ожидание от обеих частей (27) с использованием этого равенства мы получаем переходом к пределу при $n \to \infty$, что функция распределения nT_n сходится при выполнении гипотезы H_1 к распределению случайной величины

$$(aL + b)^2, \tag{28}$$

где

$$b = b_1 + b_2,$$

а то же, что и в части (1).

5. Заключение

В данной работе получено асимптотическое распределение рассматриваемого критерия и найдена формула для асимптотической мощности в случае функций g общего вида для альтернативных распределений, отличающися от нулевого величиной параметра сдвига и/или параметра масштаба. С помощью статистического 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.

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), pp. 109-119 (2005).

2. Melas V. and Salnikov D.: On Asymptotic Power of the New Test for Equality of Two Distributions. In: A. N. Shiryaev et al. (eds.), Recent Developments in Stochastic Methods and Applications, Springer Proceedings in Mathematics and Statistics 371, pp. 204-214 (2021).

3. Мелас В.Б. Об асимптотической мощности одного метода проверки гипотез о равенстве распределений. Вестник СПбГУ, Сер. 1, Вып. 2 (2023)(в печати).

4. Rocha, D.F.S., Bittencourt, I.I., de Amorim Silva, R. et al. An assistive technology based on Peirce's semiotics for the inclusive education of deaf and hearing children. Univ Access Inf Soc (2022).

5. Hoeffding W.: A class of statistics with asymptotically normal distribution. Ann. Math. Statistics 19, pp. 293-325 (1948).