УДК 519.21

Д. А. Клюшин, Ю. И. Петунин (Киев. нац. ун-т им. Т. Шевченко)

НЕПАРАМЕТРИЧЕСКИЙ КРИТЕРИЙ ЭКВИВАЛЕНТНОСТИ ГЕНЕРАЛЬНЫХ СОВОКУПНОСТЕЙ, ОСНОВАННЫЙ НА МЕРЕ БЛИЗОСТИ МЕЖДУ ВЫБОРКАМИ

We propose a new proximity measure between samples which is based on confidence limits for the bulk of general population. The confidence limits are constructed by means of order statistics. For this proximity measure, we compute approximate confidence limits corresponding to a given significance level in the cases where the null hypothesis of the equality of hypothetical distribution functions may be true as well as false. We compare considered proximity measure with the Kolmogorov – Smirnov statistics and the Wilcoxon statistics for samples from various populations. On the basis of the proposed proximity measure, we construct statistical criterion for testing hypothesis of the equality of hypothetical distribution functions.

Пропонується нова міра близькості між вибірками, яка базується на довірчих межах для основної розподіленої маси значень генеральної сукупності, побудованих за допомогою порядкових статистик. Для цієї міри близькості обчислюються наближені межі, що відповідають заданому рівню значущості у випадках, коли нульова гіпотеза про рівність гіпотетичних функцій розподілу може бути як вірною, так і хибною. Проводиться порівняння цієї міри близькості зі статистиками Колмогорова – Смирнова і Вілкоксона для вибірок із різноманітних генеральних сукупностей. На підставі запропонованої міри близькості побудовано статистичний критерій для перевірки гіпотези про рівність гіпотетичних функцій розподілу.

1. Введение. Пусть $x = (x_1, x_2, \ldots, x_n)$ и $x' = (x'_1, x'_2, \ldots, x'_m)$ — выборки из генеральных совокупностей G и G' соответственно и $z = (z_1, z_2, \ldots, z_k)$ — выборка, принадлежащая одной из этих совокупностей. Полагаем, что все выборки получены путем простого случайного выбора. Необходимо идентифицировать генеральную совокупность, из которой взята выборка z.

Эту проблему можно решать с помощью классических непараметрических критериев Колмогорова – Смирнова, Вилкоксона или любого другого непараметрического двухвыборочного критерия [1]. Однако при использовании этих критериев применяются односторонние доверительные границы, соответствующие заданному уровню значимости (например, 5-процентному), что может привести к неопределенности и непринятию решения. Действительно, предположим, что мы используем критерий Колмогорова – Смирнова при сравнении выборки z с x и x'. Для этого вычисляются статистики Колмогорова – Смирнова $\rho(z, x)$ и $\rho(z, x')$, а затем находятся уровни значимости $t_\beta(k, n)$ и $t_\beta(k, m)$. Если $\rho(z, x) \geq t_\beta(k, n)$ и $\rho(z, y) < t_\beta(k, m)$, то считается, что $z \in G'$; если же $\rho(z, x) < t_\beta(k, n)$ и $\rho(z, y) \geq t_\beta(k, m)$, то $z \in G$; в случае остальных возможных неравенств возникает неопределенность и никакое решение не принимается. Другим существенным недостатком односторонних непараметрических критериев является тот факт, что вероятность попадания статистики $\rho(z, x)$ в доверительный интервал $[0, t_\beta(k, n)]$ (т. е. его доверительный уровень) можно точно определить лишь при условии, когда $z \in G$ (гипотеза H); в противном случае эта вероятность может принимать любые значения от 0 до 1 и является неизвестной. В связи с этим даже в случае принятия решения невозможно оценить вероятности ошибок 2-го рода, возникающие при использовании этого критерия. Эта ситуация характерна для любых односторонних непараметрических критериев.

Основная цель данной работы состоит в определении меры близости (p-статистики) $\rho_p(z, x)$ между выборками z и x, для которой можно построить двусторонний доверительный интервал $\left[\rho_p^{(l)}, \rho_p^{(u)}\right]$, соответствующий заданному уровню значимости, причем этот уровень значимости не зависит от истинности или ложности гипотезы H. Кроме того, на основании этой меры близости

строится непараметрический критерий эквивалентности генеральных совокупностей, причем в некоторых практически важных случаях можно оценить его вероятность ошибки 1- и 2-го рода.

2. Доверительные интервалы и уровни значимости. Пусть $x = (x_1, \ldots, x_n)$ — выборка из генеральной совокупности G и p — некоторый известный или неизвестный показатель, значения которого могут зависеть от выборки x. Рассмотрим две непрерывные функции $a(u_1, \ldots, u_n)$ и $b(u_1, \ldots, u_n)$ от n переменных $u_1, \ldots, u_n$, удовлетворяющие неравенству

$$a(u_1, \ldots, u_n) < b(u_1, \ldots, u_n) \qquad \forall (u_1, \ldots, u_n) \in R^n.$$

Случайный интервал $(a(u_1, \ldots, u_n), b(u_1, \ldots, u_n)) = (a, b)$ называется доверительным интервалом для p, соответствующим уровню значимости β, если

$$\mathrm{P}(p \in (a, b)) = 1 - \beta, \quad 0 \le \beta \le 1;$$

при этом числа $a = a(x_1, \ldots, x_n)$, $b = b(x_1, \ldots, x_n)$ называются доверительными границами для p, соответствующими уровню значимости β. Это понятие можно обобщить на случай последовательности показателей $p_1, p_2, \ldots, p_k, \ldots$.

Определение. *Интервалы*

$$(a_k, b_k) = (a_k(x_1, \ldots, x_k), b_k(x_1, \ldots, x_k)), \quad k = 1, 2, \ldots,$$

называются асимптотическими интервалами для показателей p_i, $i = 1, 2, \ldots, k, \ldots$, *соответствующими уровню значимости* β, *если*

$$\lim_{k\to\infty} \mathrm{P}(p_k \in (a_k(x_1, \ldots, x_k), b_k(x_1, \ldots, x_k))) = 1 - \beta, \tag{1}$$

а концы этих интервалов $a_k(x_1, \ldots, x_k)$ *и* $b_k(x_1, \ldots, x_k)$ *— асимптотическими доверительными границами.*

Замечание. Для величины β будем также использовать термин „асимптотический уровень значимости последовательности (a_k, b_k)".

В частности, если все показатели $p_1, p_2, \ldots, p_k, \ldots$ одинаковы, т. е. $p_k = p$ $\forall k = 1, 2, \ldots$, то интервал (a_k, b_k) называется асимптотическим доверительным интервалом показателя p, а величина β в формуле (1) — асимптотическим уровнем значимости интервала (a_k, b_k).

Во многих практически важных частных случаях вычислить точное значение уровня значимости β_k доверительного интервала (a_k, b_k) довольно сложно, а определить асимптотический уровень значимости β, используя соответствующие предельные теоремы, достаточно просто. Если $\beta = \lim\limits_{k\to\infty} \beta_k$, то β можно принять за приближенное значение истинного уровня значимости β_k доверительных интервалов (a_k, b_k): $\beta_k \approx \beta$. В практике статистических вычислений при $n \ge 30$ величина β является хорошим приближением для β_k (см., например, [2, с. 410], табл. 7). В связи с этим будем также называть число β приближенным уровнем значимости интервалов (a_k, b_k).

3. Мера близости между выборками (p-статистика). Обозначим через H гипотезу о равенстве непрерывных функций распределения $F_G(u)$ и $F_{G'}(u)$ генеральных совокупностей G и G' соответственно. Пусть $x = (x_1, \ldots, x_n) \in G$ и $x' = (x'_1, \ldots, x'_m) \in G'$, $x_{(1)} \le \ldots \le x_{(n)}$, $x'_{(1)} \le \ldots \le x'_{(m)}$ — порядковые статистики. Предположим, что $F_G(u) = F_{G'}(u)$. Обозначим через $A_{ij}^{(k)}$, $k = 1, 2, \ldots, m$, случайное событие, состоящее в том, что x'_k попадает в интервал $\left(x_{(i)}, x_{(j)}\right)$, т. е. $A_{ij}^{(k)} = \left\{x'_k \in \left(x_{(i)}, x_{(j)}\right)\right\}$. Как известно [3], в случае, когда $F_G(u) = F_{G'}(u)$

(т. е. $G = G'$), вероятность этого события вычисляется по формуле

$$\mathrm{P}\left(A_{ij}^{(k)}\right) = \mathrm{P}\left(x'_k \in \left(x_{(i)}, x_{(j)}\right)\right) = p_{ij}^{(n)} = \frac{j-i}{n+1} = \frac{q}{n+1}, \quad q = j-i. \tag{2}$$

Положим

$$p_{ij}^{(1)} = \frac{h_{ij}^{(n)} m + g^2/2 - g\sqrt{h_{ij}^{(n)}(1-h)m + g^2/4}}{m+g^2},$$
$$p_{ij}^{(2)} = \frac{h_{ij}^{(n)} m + g^2/2 + g\sqrt{h_{ij}^{(n)}(1-h)m + g^2/4}}{m+g^2}, \tag{3}$$

где $h_{ij}^{(n)}$ — частота появления события $A_{ij}^{(n)}$ в m испытаниях. Величина g определяет уровень значимости доверительного интервала $I_{ij}^{(n,m)} = \left(p_{ij}^{(1)}, p_{ij}^{(2)}\right)$; в силу правила 3σ при $g = 3$ уровень значимости этого интервала не превышает 0,05. (Подробное описание и обоснование правила 3σ см. в [4].)

Обозначим через N $(N = n(n-1)/2)$ количество всех доверительных интервалов $I_{ij}^{(n,m)} = \left(p_{ij}^{(1)}, p_{ij}^{(2)}\right)$ и через L количество тех интервалов $I_{ij}^{(n,m)}$, которые содержат вероятности $p_{ij}^{(n)}$. Положим

$$h^{(n,m)} = \rho(F^*, F^{*\prime}) = \rho(x, x') = \frac{L}{N}.$$

Поскольку $h^{(n,m)}$ — частота случайного события $B = \left\{p_{ij}^{(n)} \in I_{ij}^{(n,m)}\right\}$, имеющего вероятность $p(B) = 1-\beta$, полагая $h_{ij}^{(n,m)} = h^{(n)}$, $m = N$ и $g = 3$ в формулах (3), получаем доверительный интервал $I^{(n,m)} = (p^{(1)}, p^{(2)})$ для вероятности $p(B)$. Статистика $h^{(n)}$ называется p-статистикой и является мерой близости $\rho(x, x')$ между выборками x и x'. Доверительные интервалы $I_{ij}^{(n,m)} = \left(p_{ij}^{(1)}, p_{ij}^{(2)}\right)$ и $I^{(n,m)} = (p^{(1)}, p^{(2)})$ будем называть интервалами, построенными по правилу 3σ.

4. Оценка асимптотического доверительного уровня значимости. В дальнейшем будем предполагать, что выборки x и x' имеют одинаковый объем, равный n. Обоснование корректности предложенной меры близости связано с проблемой вычисления уровня значимости соответствующих доверительных интервалов $I_{ij}^{(n)}$ и $I^{(n)}$. Для схемы испытаний Бернулли асимптотические уровни значимости этих интервалов при $n \to \infty$ можно вычислить на основании соответствующих предельных теорем, однако в рассматриваемом случае эти теоремы неприменимы, поскольку исследуемая схема испытаний не является схемой Бернулли (случайные события $A_{ij}^{(k)}$, $k = 1, 2, \ldots, n$, не являются независимыми). Схема испытаний, при которой могут появляться события $A_{ij}^{(k)}$, $k = 1, 2, \ldots, n$, в случае истинности гипотезы H называется обобщенной схемой Бернулли [5, 6]. Если же гипотеза H неверна, указанная схема испытаний называется модифицированной схемой Бернулли [5, 6]. В общем случае, когда может быть истинной любая гипотеза, как $F_G(u) = F_{G'}(u)$, так и $F_G(u) \neq F_{G'}(u)$, эта схема испытаний согласно терминологии С. Котца и Н. Джексона [7] называется МП-схемой.

Найдем оценку для асимптотического уровня значимости доверительных интервалов $I_{ij}^{(n)}$.

Теорема 1. *Если в обобщенной схеме испытаний Бернулли* 1) $n = m$; 2) $0 < \lim\limits_{n\to\infty} p_q^{(n)} = p_0 < 1$ *и* 3) $0 < \lim\limits_{n\to\infty} i/(n+1) = p^* < 1$, *то асимптотический уровень значимости* β *последовательности доверительных интервалов* $I_{ij}^{(n)}$ *для вероятностей* $p_q^{(n)} = p_{ij}^{(n)} = p(A_{ij}^{(n)})$, *построенных по правилу* 3σ, *не превышает* 0,05.

Доказательство. Анализ вывода формул (3) показывает, что уровень значимости β_n интервала $I_{ij}^{(n)}$ равен уровню значимости $\tilde{\beta}_n$ доверительного интервала

$$\tilde{I}_n = \left(h_{ij}^{(n)} - 3\sqrt{\frac{p_q^{(n)}\left(1-p_q^{(n)}\right)}{n}},\ h_{ij}^{(n)} + 3\sqrt{\frac{p_q^{(n)}\left(1-p_q^{(n)}\right)}{n}} \right) \tag{4}$$

для вероятности $p_q^{(n)} = p_{ij}^{(n)}$ при условии, что $\mathrm{E}(h_{ij}^{(n)}) = p_{ij}^{(n)}$, поэтому

$$\beta = \lim_{n\to\infty} \mathrm{P}\left(p_q^{(n)} \notin \tilde{I}_n\right).$$

В работе [5] показано, что в этом случае математическое ожидание частоты $h_{ij}^{(n)}$ равно $p_{ij}^{(n)}$:

$$\mathrm{E}\left(h_{ij}^{(n)}\right) = p_{ij}^{(n)} = p_q^{(n)},$$

а ее дисперсия определяется формулой

$$\mathrm{D}\left(h_{ij}^{(n)}\right) = \frac{p_{ij}^{(n)}\left(1-p_{ij}^{(n)}\right)}{n}\,\frac{2n+1}{n+2} = \alpha_n \sigma_{ij,B}^{(n)},$$

где

$$\alpha_n = \frac{2n+1}{n+2} < 2, \qquad \sigma_{ij,B}^{(n)} = \frac{p_{ij}^{(n)}\left(1-p_{ij}^{(n)}\right)}{n}.$$

Легко видеть, что доверительный уровень интервала $\tilde{I}_n$ совпадает с доверительным уровнем интервала $\hat{I}_n$ для частоты $h_{ij}^{(n)}$

$$\hat{I}_n = \left(p_q^{(n)} - 3\sigma_{ij,B}^{(n)},\ p_q^{(n)} + 3\sigma_{ij,B}^{(n)}\right).$$

Поскольку $\alpha_n < 2$, то

$$\begin{aligned}
\hat{I}_n &= \left(p_q^{(n)} - \frac{3}{\sqrt{2}}\sqrt{2}\sigma_{ij,B}^{(n)},\ p_q^{(n)} + \frac{3}{\sqrt{2}}\sqrt{2}\sigma_{ij,B}^{(n)}\right) \approx \\
&\approx \left(p_q^{(n)} - 2{,}1213\sqrt{2}\sigma_{ij,B}^{(n)},\ p_q^{(n)} + 2{,}1213\sqrt{2}\sigma_{ij,B}^{(n)}\right) \supset \\
&\supset \left(p_q^{(n)} - 2{,}1213\sigma\left(h_{ij}^{(n)}\right),\ p_q^{(n)} + 2{,}1213\sigma\left(h_{ij}^{(n)}\right)\right) \supset \\
&\supset \left(p_q^{(n)} - 1{,}96\sigma\left(h_{ij}^{(n)}\right),\ p_q^{(n)} + 1{,}96\sigma\left(h_{ij}^{(n)}\right)\right) = \bar{I}_n.
\end{aligned} \tag{5}$$

В силу условий данной теоремы, а также предельной теоремы, доказанной в [6] (теорема 10), в обобщенной схеме Бернулли случайная величина $h_{ij}^{(n)}$ является асимптотически нормальной с параметрами

$$\mathrm{E}\left(h_{ij}^{(n)}\right) = p_q^{(n)} \quad \text{и} \quad \mathrm{D}\left(h_{ij}^{(n)}\right) = \sigma^2\left(h_{ij}^{(n)}\right),$$

поэтому асимптотический уровень значимости интервала $\bar{I}_n$ равен 0,05. По-

скольку $\bar{I}_n \subset \hat{I}_n$, асимптотический уровень значимости интервала $\bar{I}_n$, а значит, и интервалов $\tilde{I}_n$ и $I_{ij}^{(n)}$ не превышает 0,05.

Теорема доказана.

Перейдем теперь к вычислению уровня значимости доверительного интервала $I^{(n)}$, построенного по правилу 3σ. Рассмотрим вначале случай, когда справедлива гипотеза H, т. е. $F_G(u) = F_{G'}(u)$.

Покажем, что эту проблему можно свести к исследованию случайных величин ξ и ξ', равномерно распределенных на отрезке $[0, 1]$. Действительно, пусть ξ — произвольная случайная величина из генеральной совокупности G со строго возрастающей непрерывной функцией распределения $F_G(u)$, $(\xi_1, \ldots, \xi_n)$ — выборка из G, а $\xi_{(1)} \le \ldots \le \xi_{(n)}$ — ее вариационный ряд. Рассмотрим случайную величину $\eta = F_G(\xi)$. Известно, что η имеет равномерное распределение $F_\eta(u)$ на отрезке $[0, 1]$. Кроме того, в силу монотонности функции распределения $F_G(u)$

$$\eta_{(i)} = F_G(\xi_{(i)}),$$

где $\eta_{(i)}$ — порядковые статистики, построенные по выборке $\eta_i = F_G(\xi_i)$. Далее, случайная величина

$$\eta' = F_G(\eta') = F_{G'}(\eta')$$

также имеет равномерное распределение на отрезке $[0, 1]$, причем случайные события $\xi'_k \in (\xi_{(i)}, \xi_{(j)})$ и $\eta'_k \in (\eta_{(i)}, \eta_{(j)})$ эквивалентны, поскольку функция $F_G(u)$ является монотонно возрастающей. Отсюда следует, что частота $h_{ij}^{(\xi)}$ случайного события, состоящего в том, что $\{\xi'_k \in (\xi_{(i)}, \xi_{(j)})\}$, совпадает с частотой $h_{ij}^{(\eta)}$ случайного события $\{\eta'_k \in (\eta_{(i)}, \eta_{(j)})\}$, поэтому эквивалентны между собой события

$$\{p_{ij} \in (p_{ij}^{(1)}(\xi), p_{ij}^{(2)}(\xi))\} \quad \text{и} \quad \{p_{ij} \in (p_{ij}^{(1)}(\eta), p_{ij}^{(2)}(\eta))\},$$

где $(p_{ij}^{(1)}(\xi), p_{ij}^{(2)}(\xi))$ и $(p_{ij}^{(1)}(\eta), p_{ij}^{(2)}(\eta))$ — доверительные границы для p_{ij}, построенные с помощью формул (3) при $g = 3$ на основании частот $h_{ij}^{(\xi)}$ и $h_{ij}^{(\eta)}$. Таким образом, мера близости между выборками $(\xi_1, \ldots, \xi_n)$ и $(\xi'_1, \ldots, \xi'_n)$ совпадает с мерой близости между выборками $(\eta_1, \ldots, \eta_n)$ и $(\eta'_1, \ldots, \eta'_n)$. Следовательно, уровень значимости доверительного интервала $I^{(n)} = (p^{(1)}, p^{(2)})$ для меры близости можно вычислить с помощью машинного моделирования, используя для этого выборки случайных величин, равномерно распределенных на отрезке $[0, 1]$. Результаты машинного моделирования показывают, что для выборок объема $n = 100$ среднее значение доверительного уровня интервала $I^{(100)}$, вычисленное для 30 выборок, равно 0,997, что превышает, как и следовало ожидать на основании теоремы 1, величину 0,95. Кроме того, не было отмечено ни одного случая выхода числа 0,95 за пределы интервала $I^{(100)}$.

Перейдем теперь к оценке уровня значимости доверительного интервала $I^{(n)} = (p^{(1)}, p^{(2)})$ в случае, когда гипотеза H неверна (т. е. $F_G(u) \ne F_{G'}(u)$).

Напомним, что случайным экспериментом называется последовательность случайных испытаний $T_1, T_2, \ldots, T_n, \ldots$, представляющих собой повторение одного и того же случайного базового испытания T [8], а вероятность $p(A)$ события A, которое может произойти при реализации эксперимента, определяет-

ся как предел частоты $h_n(A)$, когда количество испытаний n стремится к бесконечности: $p(A) = \lim\limits_{n\to\infty} h_n(A)$ [8–10].

Пусть $B_1, B_2, \ldots, B_n, \ldots$ — последовательность событий, которые могут осуществиться при проведении случайного эксперимента, и $\lim\limits_{n\to\infty} p(B_k) = p^*$. Пусть $h_{n_1}(B_1), h_{n_2}(B_2), \ldots, h_{n_k}(B_k), \ldots$ — последовательность частот событий B_1, $B_2, \ldots, B_n, \ldots$ соответственно, причем $k/n_k \to 0$ при $k \to \infty$. Будем называть эксперимент *усиленным случайным экспериментом*, если $h_{n_k}(B_k) \to p^*$ при $k \to \infty$. Отметим, что в данном случае базовым испытанием T является простой случайный выбор одного элемента x из генеральных совокупностей G и G'.

Теорема 2. *В условиях усиленного случайного эксперимента, если выборки* $x = (x_1, \ldots, x_n) \in G$ *и* $x' = (x'_1, \ldots, x'_m) \in G'$ *имеют одинаковый объем, асимптотический уровень значимости интервала* $I^{(n)} = (p^{(1)}, p^{(2)})$, *построенный по правилу* 3σ *при* $g = 3$ *с помощью формул* (3), *не превышает* 0,05.

Доказательство. Рассмотрим доверительные интервалы $I_{ij}^{(n)} = \left(p_{ij}^{(1)}, p_{ij}^{(2)}\right)$, $i < j$, построенные по частоте $h_{ij}^{(n)}$ случайного события $\left\{x'_k \in \left(x_{(i)}, x_{(j)}\right)\right\}$ с помощью правила 3σ. Поскольку $F_G(u) \neq F_{G'}(u)$, вероятность $\mathrm{P}\left(B_{ij}^{(n)}\right)$ случайного события

$$B_{ij}^{(n)} = \left\{p_{ij}^{(n)} = \frac{j-i}{n+1} \in I_{ij}^{(n)}\right\},$$

вообще говоря, не равна 0,95 и может зависеть от индексов i и j. Введем случайные величины

$$\delta_{ij}^{(n)} = \begin{cases} 1, & \text{если} \quad p_{ij}^{(n)} \in I_{ij}^{(n)}; \\ 0, & \text{если} \quad p_{ij}^{(n)} \notin I_{ij}^{(n)}, \end{cases}$$

$$z^{(n)} = \frac{1}{N}\sum_{i<j} \delta_{ij}^{(n)} = h^{(n)}, \qquad N = \frac{n(n-1)}{2}.$$

Тогда

$$p^{(n)} = \mathrm{E}\left(z^{(n)}\right) = \frac{1}{N}\sum_{i<j} p\left(B_{ij}^{(n)}\right). \tag{6}$$

Пронумеруем случайные величины $\delta_{ij}^{(n)}$ произвольным образом и обозначим полученное множество через $X = (X_1, X_2, \ldots, X_N)$, а совместную функцию распределения координат случайного вектора X — через $F(u_1, u_2, \ldots, u_N)$. В работе [11] предложен следующий метод получения выборки из множества X: выбираем произвольно один из элементов множества X и обозначаем его через γ_1; затем из множества $X \setminus \{\gamma_1\}$ оставшихся элементов выбираем произвольно элемент γ_2 и так далее. Таким образом, мы имеем серию испытаний, состоящую из N случайных выборов без возвращения элементов множества X, содержащихся в урне. Получившуюся при этом многомерную случайную величину $\gamma^{(N)} = (\gamma_1, \gamma_2, \ldots, \gamma_N)$ будем называть индуцированной выборкой, полученной с помощью урновой модели. Как известно [12], функция совместного распределения $F_{\gamma^{(N)}}(u_1, u_2, \ldots, u_N)$ индуцированной выборки $\gamma^{(N)}$ имеет вид

$$F_{\gamma^{(N)}}(u_1, u_2, \ldots, u_N) = \frac{1}{N!} \sum_{(i_1, i_2, \ldots, i_N) \in G_N} F\left(u_{i_1}, u_{i_2}, \ldots, u_{i_N}\right), \tag{7}$$

где G_N — группа перестановок чисел $(1, 2, \ldots, N)$. Из формулы (6) следует, что $F_{\gamma^{(N)}}(u_1, u_2, \ldots, u_N)$ является симметрической функцией, так что элементы γ_k, $k = 1, 2, \ldots, N$, индуцированной выборки являются симметрично зависимыми случайными величинами с маргинальной функцией распределения

$$F_{\gamma_k}(u) = \frac{1}{N!}[F(u, \infty, \ldots, \infty)(N-1)! + F(\infty, u, \infty, \ldots, \infty) \times$$

$$\times (N-1)! + \ldots + F(\infty, \ldots, \infty, u)(N-1)!] = \frac{1}{N}[F_1(u) + \ldots + F_N(u)].$$

Отсюда следует, что все маргинальные функции распределения $F_{\gamma_k}(u)$ одинаковы и имеют вид

$$F_{\gamma_k}(u) = \begin{cases} 0, & \text{если} \quad u \le 0; \\ q^{(n)} = 1 - p^{(n)}, & \text{если} \quad 0 < u < 1; \\ 1, & \text{если} \quad u \ge 1, \end{cases}$$

где $p^{(n)} = \mathrm{E}(z)$ определяется по формуле (6). Обозначим через B_n случайное событие, состоящее в том, что вероятность $p_{ij}^{(n)} = (j-1)/(n+1)$ принадлежит доверительному интервалу $I_{ij}^{(n)}$ при случайном выборе этого интервала (или вероятности $p_{ij}^{(n)}$). Очевидно, что событие B_n происходит тогда и только тогда, когда случайно выбранные $\delta_{ij}^{(n)} = \gamma_l = 1$, $l = 1, 2, \ldots, N$, поэтому индуцированную выборку $\gamma^{(N)}$ можно интерпретировать как схему испытаний, в результате которых событие B_n может появиться с вероятностью $p(B_n) = p^{(n)}$. Легко видеть, что частота h_n этого события совпадает с $h = \rho(x, x')$, так что

$$h = \frac{1}{N} \sum_{k=1}^{N} \gamma_k$$

и

$$\mathrm{E}(h) = \frac{1}{N} \sum_{k=1}^{N} \mathrm{E}(\gamma_k) = p^{(n)},$$

$$\mathrm{D}(h) = \frac{1}{N^2} \left(\sum_{k=1}^{N} \mathrm{D}(\gamma_k) + \sum_{k \ne s} K(\gamma_k, \gamma_s) \right). \tag{8}$$

Очевидно, что $\mathrm{D}(\gamma_k) = p^{(n)} q^{(n)}$, а коэффициент ковариации $K(\gamma_k, \gamma_s)$ не зависит от k и s: $K(\gamma_k, \gamma_s) \equiv C_N$. Действительно, при любых k и s двумерная функция распределения γ_k и γ_s не зависит от k и s, так как в силу симметричности функций $F_{\gamma^{(N)}}(u_1, \ldots, u_N)$ имеем

$$F_{\gamma_k \gamma_s}(u_k, u_s) = F_{\gamma^{(N)}}(\infty, \ldots, \infty, u_k, \infty, \ldots, \infty, u_s, \infty, \ldots, \infty) =$$

$$= F_{\gamma^{(N)}}(u_k, u_s, \infty, \ldots, \infty) = F_0(u_k, u_s).$$

Следовательно, коэффициент ковариации $K(\gamma_k, \gamma_s)$, который определяется

функцией $F_{\gamma_k\gamma_s}(u_k, u_s)$, принимает постоянное значение при фиксированном N и не зависит от γ_k и γ_s. Таким образом,

$$\frac{1}{N^2}\sum_{k\neq s} K(\gamma_k, \gamma_s) = \frac{N-1}{N} C_N.$$

Далее,

$$C_N = K(\gamma_k, \gamma_s) = \mathrm{E}(\gamma_k\gamma_s) - \left(p^{(n)}\right)^2,$$

так что смешанный момент второго порядка $\mathrm{E}(\gamma_k\gamma_s)$ тоже является постоянной величиной при фиксированном N. Поскольку величины C_N и $p^{(n)}$, входящие в формулу (8), неизвестны, заменим их оценками

$$p^{(n)} \approx h,$$

$$C_N \approx \frac{1}{N(N-1)}\sum_{k\neq s}\gamma_k\gamma_s - h^2 = \frac{L(L-1)}{N(N-1)} - \frac{L^2}{N^2} = \frac{L(L-N)}{N^2(N-1)} = Q_N,$$

причем

$$|Q_N| = h\frac{N-L}{N(N-1)}.$$

Такая замена является корректной, так как согласно определению асимптотического уровня значимости β можно считать без ограничения общности, что вероятности $p^{(n)} = p(B_n)$ сходятся к некоторой величине p^*, а из условий теоремы следует, что $h_n \to p^*$ при $n \to \infty$, поэтому $h_n = h$ можно считать приближенным значением p_n и p^*.

Покажем теперь, что $C_N - Q_N$ сходится в среднем к нулю при $N \to \infty$.

Введем следующие обозначения:

$$\tilde{Q}_N = \frac{1}{N(N-1)}\sum_{k\neq i}\gamma_k\gamma_i - p^2,$$

$$Q_N^* = \frac{L^2}{N^2} - p^2 = h^2 - p^2,$$

тогда

$$|Q_N - C_N| \le \left|Q_N - \tilde{Q}_N\right| + \left|\tilde{Q}_N - C_N\right| = \left|h^2 - p^2\right| + \left|\tilde{Q}_N - C_N\right| \le$$
$$\le |h+p||h-p| + \left|\tilde{Q}_N - C_N\right| \le 2|h-p| + \left|\tilde{Q}_N - C_N\right|.$$

Пусть ε — произвольное сколь угодно малое положительное число. Выберем n настолько большим, чтобы

$$\left|h_n - p^{(n)}\right| = |h-p| < \frac{\varepsilon}{8} \quad \text{и} \quad \frac{1}{N} \le \frac{\varepsilon}{4}.$$

Тогда

$$m(|Q_N - C_N|) \le 2m(|h-p| + m(|\tilde{Q}_N - C_N|)) < \frac{\varepsilon}{4} + m(|\tilde{Q}_N - C_N|).$$

Без ограничения общности можно считать, что $C_N \ge 0$, ибо в противном случае

$$\mathrm{D}(h) \le \frac{1}{N^2}\sum_{k=1}^{n}\mathrm{D}(\gamma_k) = \frac{p^{(n)}q^{(n)}}{N}.$$

Отсюда следует

$$\tilde{Q}_N = \frac{L(L-1)}{N(N-1)} - p^2 \le h^2 - p^2$$

и

$$0 \le C_N = m\left(\tilde{Q}_N\right) \le m(h^2 - p^2) = m[(h-p)(h+p)],$$

$$|C_N| = C_N \le |m[(h-p)(h+p)]| \le m[|h-p||h+p|] \le 2m(|h-p|) < \frac{\varepsilon}{4}.$$

Следовательно, $0 \le C_N \le \varepsilon/4$. Далее, легко видеть, что $\tilde{Q}_N \le Q_N^*$ и

$$\left|\tilde{Q}_N - Q_N^*\right| = Q_N^* - \tilde{Q}_N = \frac{L}{N}\left(\frac{L}{N} - \frac{L-1}{N-1}\right) < \frac{h}{N} < \frac{1}{N}.$$

Поэтому

$$m\left(\left|Q_N^*\right|\right) = m\left(\left|h^2 - p^2\right|\right) < \frac{\varepsilon}{4} \quad \text{и} \quad m\left(\left|\tilde{Q}_N - Q_N^*\right|\right) < \frac{1}{N}.$$

Значит,

$$m\left(\left|\tilde{Q}_N - C_N\right|\right) = m\left(\left|\tilde{Q}_N - Q_N^* + Q_N^* - C_N\right|\right) \le$$

$$\le m\left(\left|\tilde{Q}_N - Q_N^*\right|\right) + m\left(\left|Q_N^*\right|\right) + C_N \le \frac{1}{N} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} < \frac{3\varepsilon}{4}.$$

Таким образом,

$$m\left(\left|Q_N - C_N\right|\right) \le \frac{\varepsilon}{4} + \frac{3\varepsilon}{4} = \varepsilon,$$

так что Q_N является асимптотическим приближением для C_N.

Кроме того, выборочное среднее является оптимальной оценкой в классе линейных несмещенных оценок для симметрично зависимых случайных величин [11].

Следовательно,

$$\mathrm{D}(h) = \frac{p^{(n)}\left(1-p^{(n)}\right)}{N} + \frac{N-1}{N}C_N \le \frac{p^{(n)}\left(1-p^{(n)}\right)}{N} + \frac{N-1}{N}|C_N| \approx \frac{h(1-h)}{N} +$$

$$+ \frac{N-1}{N}|Q| = \frac{h(1-h)}{N} + \frac{h}{N}\frac{N-L}{N} = \frac{h(1-h)}{N} + \frac{h(1-h)}{N} = 2\frac{h(1-h)}{N},$$

так что

$$\mathrm{D}(h) \le 2\frac{h(1-h)}{N}.$$

Используя вложения интервалов (5) и центральную предельную теорему для суммы симметрично зависимых случайных величин [13] (гл. VIII), получаем, что асимптотический уровень значимости доверительного интервала $I^{(n)}$ не превышает 0,05.

Теорема доказана.

5. Критерий проверки гипотезы о равенстве функций распределения. Критерий T проверки гипотезы H о равенстве непрерывных гипотетических функций распределения $F_G(u)$ и $F_{G'}(u)$ с уровнем значимости, приближенно равным 0,05, можно сформулировать следующим образом: если доверительный интервал $I^{(n)} = (p^{(1)}, p^{(2)})$ содержит вероятность $p(B) = 1 - \beta = 0{,}95$, то гипотеза H принимается, в противном случае она отвергается и считается справедливой альтернативная гипотеза.

Одной из характерных особенностей p-статистики в отличие от других мер близости является тот факт, что уровень значимости доверительного интерва-

ла $I^{(n)}$, содержащего вероятность $p^{(n)}$, определяемую формулой (6), не превышает 0,05, когда гипотеза H неверна (теорема 2). Это дает возможность оценить в некоторых практически важных частных случаях вероятность ошибки второго рода, когда осуществляется проверка двух конкурирующих альтернативных гипотез с помощью статистических критериев, использующих p-статистики. Пусть $x = (x_1, x_2, \dots, x_n)$ и $x' = (x'_1, x'_2, \dots, x'_n)$ — выборки, элементы которых взяты из генеральных совокупностей G и G' соответственно, и $z = (z_1, z_2, \dots, z_n)$ — выборка, принадлежащая одной из этих совокупностей. Гипотеза H_0 означает, что элементы z взяты из G, а $\overline{H}_0$ — из G'; тогда H_0 и $\overline{H}_0$ являются конкурирующими альтернативными гипотезами. Обозначим через $h_{zx}(h_{zx'})$ p-статистику для выборок z и x (z и x'), а через $I_{zx} = \left(p_{zx}^{(1)}, p_{zx}^{(2)}\right)$ $\left(I_{zx'} = \left(p_{zx'}^{(1)}, p_{zx'}^{(2)}\right)\right)$ доверительный интервал, о котором идет речь в критерии T, когда параметр g, используемый для построения интервала I_3 из теоремы 1, равен 3. Приведем критерий проверки истинности гипотез H_0 и $\overline{H}_0$:

1) если $p(B) = 0{,}95 \in I_{zx}$ и $p_{zx'}^{(2)} < p_{zx}^{(1)}$, то принимается гипотеза H_0, а $\overline{H}_0$ отклоняется;

2) если $p(\overline{B}) = 0{,}95 \in I_{zx'}$ и $p_{zx}^{(2)} < p_{zx'}^{(1)}$, то принимается гипотеза $\overline{H}_0$, а H_0 отклоняется;

3) в противном случае решение об истинности гипотез H_0 и $\overline{H}_0$ не принимается.

Легко видеть, что вероятности ошибок 1- и 2-го рода этого критерия не превышают 0,05.

6. Сравнение p-статистики со статистиками Колмогорова – Смирнова и Вилкоксона. Для исследования эффективности p-статистики сравним ее с наиболее распространенными статистиками (статистикой Колмогорова – Смирнова и статистикой Вилкоксона), которые используются при построении порядковых критериев, а также при вычислении меры близости между выборками из некоторых параметрических семейств распределений. Прежде всего необходимо унифицировать эти статистики так, чтобы они принимали значения из отрезка $[0, 1]$, причем их малым значениям должны соответствовать близкие выборки.

Для статистики Колмогорова – Смирнова

$$d_{KS}(x, x') = \max_{t \in (-\infty, +\infty)} \left| F_x^*(t) - F_{x'}^*(t) \right|,$$

где $F_x^*(t)$ и $F_{x'}^*(t)$ — эмпирические функции распределения, построенные по выборкам x и x' соответственно, в такой унификации нет необходимости, поскольку статистика Колмогорова – Смирнова принимает значения из отрезка $[0, 1]$ и ее малым значениям соответствуют близкие выборки.

Для U-статистики Вилкоксона такая унификация необходима, поскольку она принимает целые неотрицательные значения. Напомним определение статистики Вилкоксона: пусть $x = (x_1, x_2, \dots, x_g)$ и $x' = (x'_1, x'_2, \dots, x'_h)$ — выборки, полученные путем простого случайного выбора из генеральных совокупностей G и G' соответственно. Определим случайные величины

$$z_{ik} = \begin{cases} 1, & \text{если} \quad x_i > x'_k; \\ 0, & \text{если} \quad x_i \le x'_k, \end{cases}$$

и

$$U = \sum_{i=1}^{g} \sum_{k=1}^{h} z_{ik}.$$

Положим $\hat{U} = gh/2$ и $J = U - \hat{U}$. Для унификации статистики Вилкоксона введем случайную величину

$$d_U(x, x') = \frac{2}{gh}|J|,$$

которую будем называть унифицированной статистикой Вилкоксона и использовать в качестве меры близости между выборками x и x'. Легко видеть, что $0 \le d_U(x, x') \le 1$ и для выборок из одной генеральной совокупности $G = G'$ статистика $d_U(x, x')$ принимает значения, близкие к нулю.

Унификация p-статистики проводится по формуле

$$d_p(x, x') = 1 - \rho(x, x').$$

Для сравнения эффективности указанных статистик были рассмотрены выборки из генеральных совокупностей, которые имеют одинаковую дисперсию при различных математических ожиданиях, одинаковое математическое ожидание при различных дисперсиях, а также выборки из генеральных совокупностей, отличающихся как математическими ожиданиями, так и дисперсиями, и выборки из генеральных совокупностей, имеющих одинаковые математические ожидания и дисперсии. Сравнивались выборки из параметрического семейства генеральных совокупностей G_α, имеющих гипотетические функции распределения $F_\alpha(u)$, $0 \le \alpha \le 1$, с выборкой из генеральной совокупности G_1 с распределением $F_1(u)$. Показателем эффективности статистики (меры близости) является значение α_0 параметра α, начиная с которого критерий, основанный на этой статистике, не различает выборки (генеральные совокупности). Введем следующие обозначения: $N(a, b)$ — нормальное распределение с математическим ожиданием a и дисперсией $b > 0$, $U(a, b)$ — равномерное распределение на отрезке $[a, b]$.

Рассмотрены следующие модели:

1. Одинаковые дисперсии при различных математических ожиданиях (рис. 1): $N(\alpha, 1)$ и $N(1, 1)$.

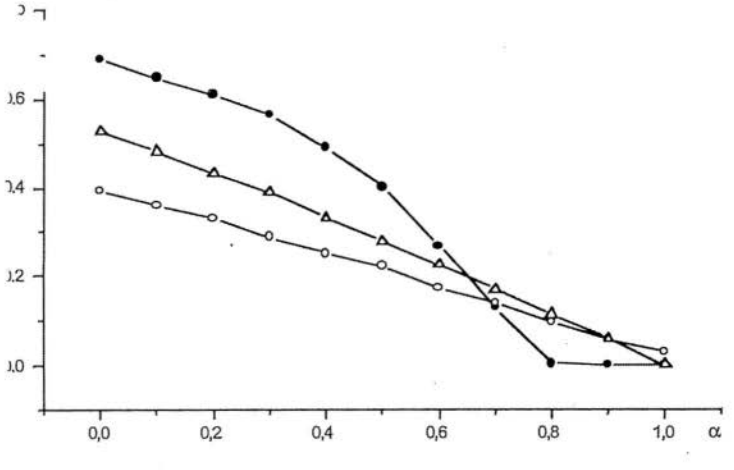

Рис. 1. Расстояния между выборкой из $N(1, 1)$ и выборками из $N(\alpha, 1)$:
—●— — PK-расстояние; —○— — KS-расстояние;
—△— — U-расстояние.

2. Одинаковые математические ожидания при различных дисперсиях ис. 2): $N(0, \alpha)$ и $N(0, 1)$.

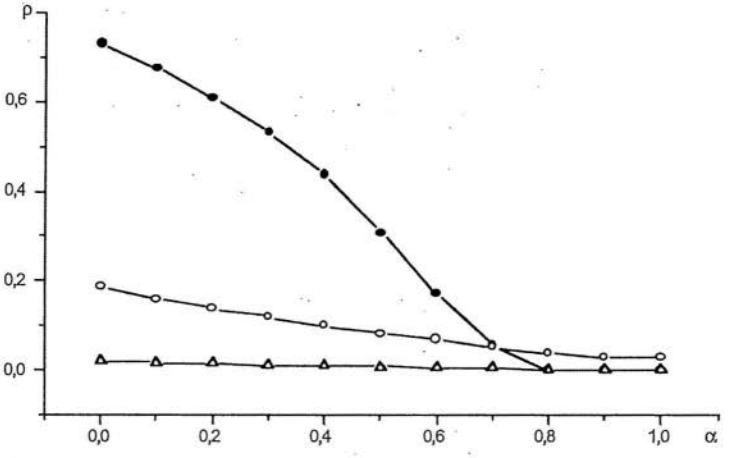

Рис. 2. Расстояния между выборкой из $N(0, 1)$ и выборками из $N(0, \alpha)$:
—●— — PK-расстояние; —○— — KS-расстояние;
—△— — U-расстояние.

3. Разные математические ожидания и дисперсии:
а) смеси распределений:

$$\alpha U(0,1) + (1-\alpha)\, U\left(\frac{1}{2}, \frac{3}{2}\right) \quad \text{и} \quad U(0,1) \quad \text{(рис. 3)},$$

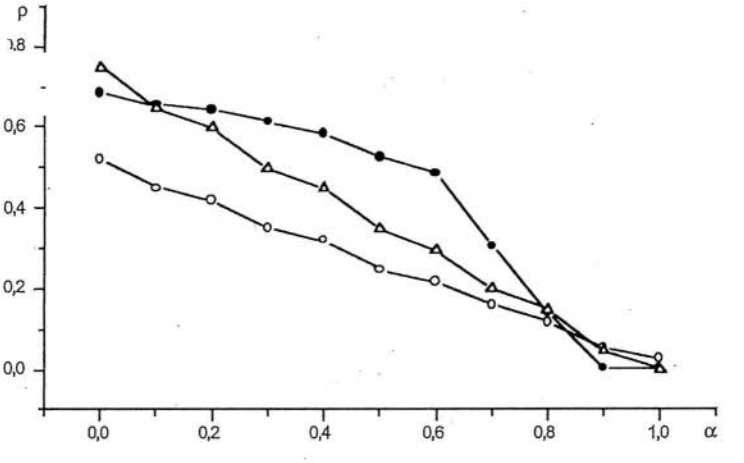

Рис. 3. Расстояния между выборкой из $U(0, 1)$ и выборками из смеси равномерных распределений $\alpha U(0, \alpha) + (1-\alpha)U(0{,}5; 1{,}5)$:
—●— — PK-расстояние; —○— — KS-расстояние;
—△— — U-расстояние.

$$\alpha U(0,1) + (1-\alpha)\, U\left(\frac{1}{6}, \frac{1}{2}\right) \quad \text{и} \quad U(0,1) \quad \text{(рис. 4)},$$

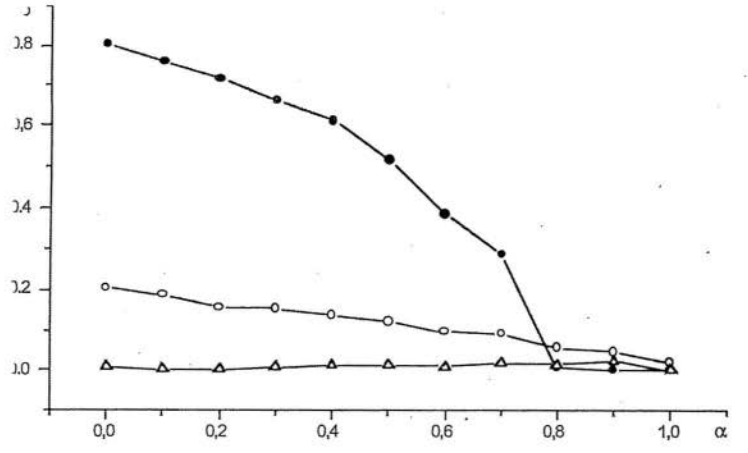

Рис. 4. Расстояния между выборкой из $U(0, 1)$ и выборками из смеси равномерного и нормального распределений $\alpha U(0, \alpha) + (1-\alpha)N(1/2$

—●— — PK-расстояние; —○— — KS-расстояние;

—△— — U-расстояние.

$\alpha N(0, 1) + (1-\alpha)N(1, 1)$ и $N(0, 1)$ (рис. 5);

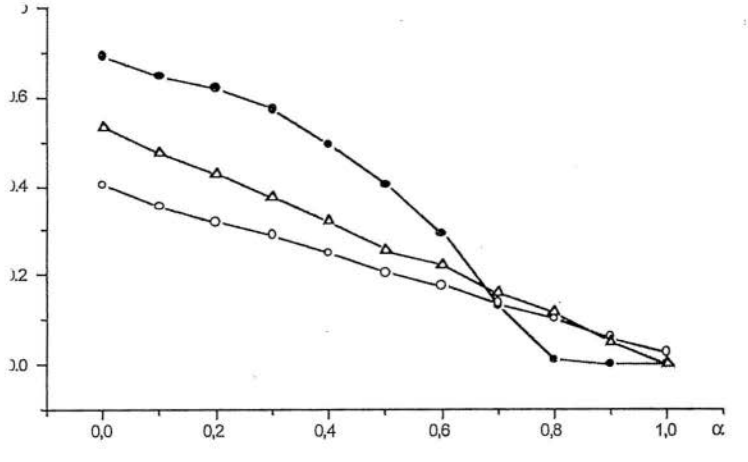

Рис. 5. Расстояния между выборкой из $N(0, 1)$ и выборками из смеси $\alpha N(0, 1) + (1-\alpha)N(1, 1)$:

—●— — PK-расстояние; —○— — KS-расстояние;

—△— — U-расстояние.

б) распределения одинакового типа:

$N(\alpha, 1+\alpha)$ и $N(1, 2)$ (рис. 6).

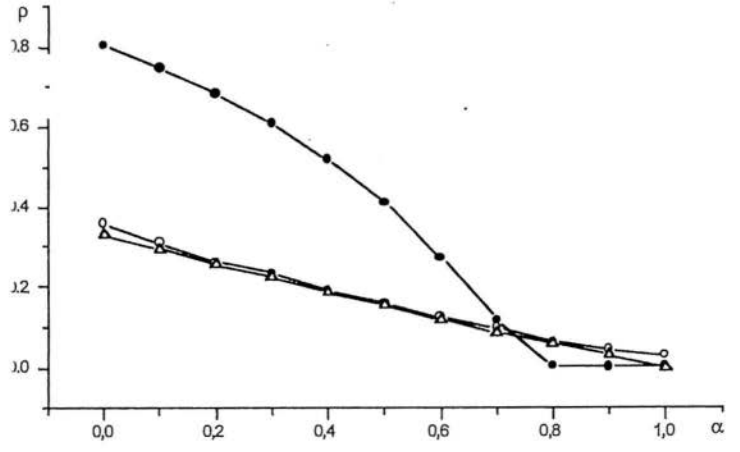

Рис. 6. Расстояния между выборкой из $N(1, 2)$ и выборками из $N(\alpha, 1+\alpha)$:

—●— — PK-расстояние; —○— — KS-расстояние;

—△— — U-расстояние.

. Альтернативы Лемана на основе распределения $U(0, 1)$:

$$f_\alpha(u) = (1+\alpha)F^\alpha(u)f(u), \quad -\frac{1}{2} \le \alpha \le 0 \quad \text{и} \quad U(0,1) \quad \text{(рис. 7)},$$

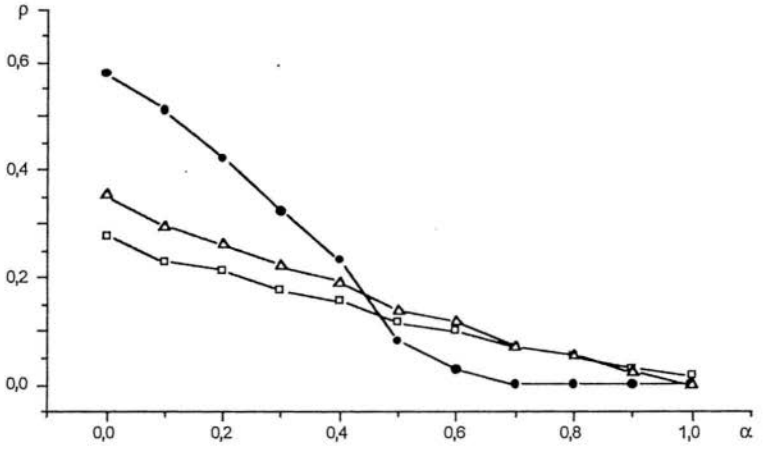

Рис. 7. Расстояния между выборкой из распределения $U(0, 1)$ и выборками из семейства альтернатив Лемана на основе распределения $U(0, 1)$ $(-0{,}5 \le u \le 0)$:

—●— — PK-расстояние; —□— — KS-расстояние;

—△— — U-расстояние.

$$f_\alpha(u) = (1+\alpha)F^\alpha(u)f(u), \quad -1 \le \alpha \le 0 \quad \text{и} \quad U(0,1) \quad \text{(рис. 8)},$$

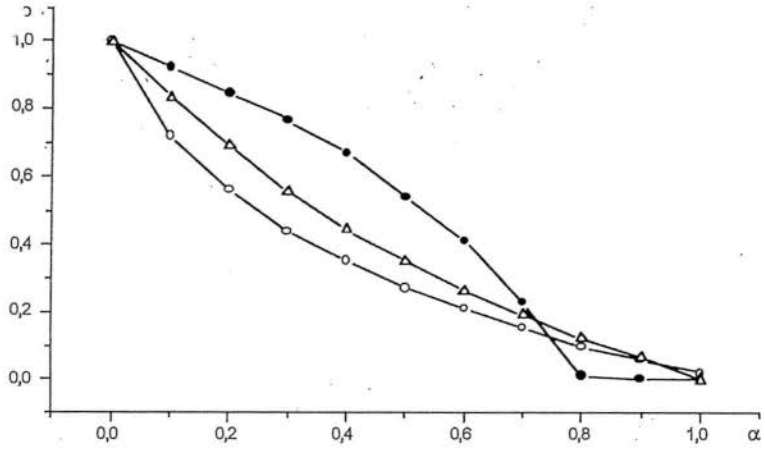

Рис. 8. Расстояния между выборкой из распределения $U(0, 1)$ и выборками из семейства альтернатив Лемана на основе распределения $U(0, 1)$ $(-1 < u \le 0)$:

—●— — PK-расстояние; —○— — KS-расстояние;

—△— — U-расстояние.

где $f(u)$ — плотность вероятностей равномерного распределения на отрезке $[0, 1]$, $f_\alpha(u)$ — плотность вероятностей альтернативы Лемана.

5. Одинаковые математические ожидания и дисперсии:

смесь распределений $\alpha N(0, 1) + (1-\alpha)\, U\left(-\sqrt[3]{3/2},\ \sqrt[3]{3/2}\right)$ и $N(0, 1)$ (рис. 9).

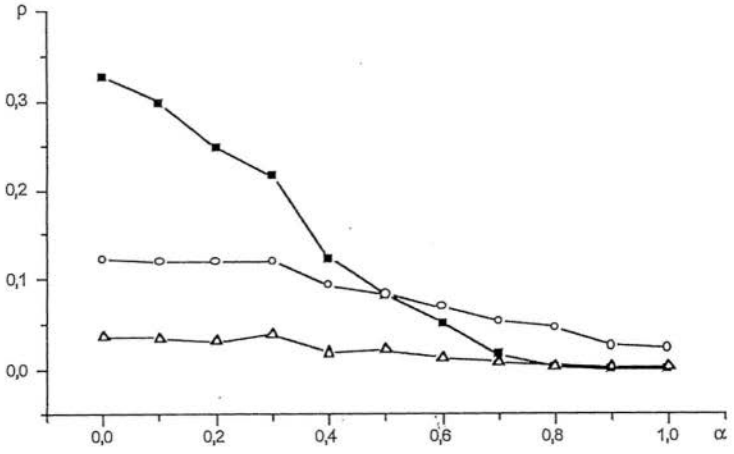

Рис. 9. Расстояния между выборкой из $N(0, 1)$ и выборками из смеси равномерного и нормального распределений $\alpha N(0, 1) + (1-\alpha)\, U\left(-\sqrt[3]{3/2},\ \sqrt[3]{3/2}\right)$ при $g = 3$:

—■— — PK-расстояние; —○— — KS-расстояние;

—△— — U-расстояние.

На рис. 1 – 9 по оси абсцисс отложены значения параметра α, а по оси ординат — значения ρ статистик, где PK-расстояние — унифицированная p-статистика, KS-расстояние — статистика Колмогорова – Смирнова, U-расстояние

— унифицированная статистика Вилкоксона. Объем каждой выборки равен 300.

Из этих рисунков видно, что в области значимых отклонений p-статистика имеет более высокую эффективность по сравнению с расстоянием Колмогорова – Смирнова и U-расстоянием. В связи с этим представляет интерес вычисление значения параметра α, начиная с которого наблюдаются значимые различия между выборками из генеральной совокупности с предельным распределением (при $\alpha = 1$) и выборками из генеральных совокупностей с распределениями соответствующими параметрам α из интервала $[0, 1)$. Эти значения α будем называть *порогом чувствительности критерия*.

В таблице приведены значения порога чувствительности для рассмотренных выше критериев и генеральных совокупностей при 5-процентном уровне значимости. Эти результаты свидетельствуют о преимуществе PK-критерия по сравнению с критериями Колмогорова – Смирнова и Вилкоксона. Однако для PK-критерия наблюдается слабая чувствительность при значениях параметра α, близких к единице (т. е. когда выборки взяты из очень близких генеральных совокупностей), в то время как у критериев Колмогорова – Смирнова и Вилкоксона этот недостаток не так сильно выражен.

Генеральная совокупность		Порог чувствительности критерия		
G_x	G_y	U	KS	PK
$N(\alpha, 1)$	$N(1, 1)$	0,9	0,7	0,8
$N(0, \alpha)$	$N(0, 1)$	—	0,2	0,7
$\alpha U(0, 1) + (1-\alpha) U\left(\frac{1}{2}, \frac{3}{2}\right)$	$U(0, 1)$	0,9	0,7	0,9
$\alpha U(0, 1) + (1-\alpha) U\left(\frac{1}{6}, \frac{1}{2}\right)$	$U(0, 1)$	—	0,2	0,8
$\alpha N(0, 1) + (1-\alpha) N(1, 1)$	$N(0, 1)$	0,9	0,5	0,8
$N(\alpha, 1+\alpha)$	$N(1, 2)$	0,8	0,5	0,8
$f_\alpha(u) = (1+\alpha) F^\alpha(u) f(u)$, $-\frac{1}{2} \le \alpha \le 0$	$U(0, 1)$	0,8	0,4	0,7
$f_\alpha(u) = (1+\alpha) F^\alpha(u) f(u)$, $-1 \le \alpha \le 0$	$U(0, 1)$	0,9	0,7	0,9
$\alpha N(0, 1) + (1-\alpha) U\left(-\sqrt[3]{\frac{3}{2}}, \sqrt[3]{\frac{3}{2}}\right)$	$N(0, 1)$	—	—	0,7

Примечание. Прочерк означает отсутствие значимых отклонений при всех значениях параметра $0 \le \alpha < 1$, так что в этом случае критерий не различает выборки.

6. Выводы. Предложенная мера близости имеет следующие преимущества по сравнению со статистиками Колмогорова – Смирнова и Вилкоксона:

1) для p-статистики можно построить приближенные доверительные границы, соответствующие заданному уровню значимости в случае, когда нулевая гипотеза H не является верной;

2) p-статистика более эффективна при проверке гипотезы об эквивалентности генеральных совокупностей с одинаковыми или близкими математическими ожиданиями (рис. 2, 4 и 9);

3) p-статистика имеет высокий порог чувствительности во всех проанализированных примерах выборок из разных генеральных совокупностей.

1. *Кокс Д., Хинкли Д.* Теоретическая статистика. – М.: Мир, 1978. – 560 с.
2. *Ван дер Варден Б. Л.* Математическая статистика. – М.: Изд-во иностр. лит., 1960. – 436 с.
3. *Мадреимов И., Петунин Ю. И.* Характеризация равномерного распределения с помощью порядковых статистик // Теория вероятностей и мат. статистика. – 1982. – Вып. 27. – С. 96 – 102.
4. *Петунин Ю. И.* Приложение теории случайных процессов в биологии и медицине. – Киев: Наук. думка, 1981. – 320 с.
5. *Матвейчук С. А., Петунин Ю. И.* Обобщение схемы Бернулли, возникающее в вариационной статистике. I // Укр. мат. журн. – 1991. – **43**, № 4. – С. 518 – 528.
6. *Матвейчук С. А., Петунин Ю. И.* Обобщение схемы Бернулли, возникающее в вариационной статистике. II // Там же. – № 6. – С. 779 – 785.
7. *Johnson N., Kotz S.* Some generalizations of Bernoulli and Polya – Eggenberger contagion models // Statist. Paper. – 1991. – **32**. – P. 1 – 17.
8. *Petunin Yu. I., Klyushin D. A.* Structure approach to solution of sixth Hilbert problem // Abstrs Int. Conf. „Functional Methods in Approximation Theory, Operator Theory, Stochastic Analysis and Statistics". – Kyiv, 2001. – P. 60.
9. *Mises R.* Wahrscheinlichkeit, Statistik und Wahrheit. – Wien, 1936.
10. *Mises R.* On the foundations of probability and statistics. – Providence: Amer. Math. Soc., 1947. – **12**. – 191 p.
11. *Курицын Ю. Г., Петунин Ю. И.* К теории линейных оценок математического ожидания случайного процесса // Теория вероятностей и мат. статистика. – 1970. – Вып. 3. – С. 80 – 92.
12. *Петунин Ю. И., Семейко Н. Г.* Случайные точечные процессы с независимым маркированием // Докл. АН СССР. – 1986. – **288**, №4. – С. 823 – 827.
13. *Феллер В.* Введение в теорию вероятностей и ее применения: В 2 т. – М.: Мир, 1967. – Т. 2. – 752 с.

Получено 30.10.2000,
после доработки — 23. 05. 2002