

сообщения
объединенного
института
ядерных
исследований
дубна

3 895

P10-88-321

П.В.Зрелов, В.В.Иванов

ПРОВЕРОЧНАЯ СТАТИСТИКА

$$w_n^{3/2} \int_{-\infty}^{\infty} [S_n(x) - P(x)]^3 dP(x)$$

В ТЕОРИИ НЕПАРАМЕТРИЧЕСКИХ
КРИТЕРИЕВ СОГЛАСИЯ

1988

ВВЕДЕНИЕ

В критериях согласия, основанных на сравнении функции распределения F , предсказываемой гипотезой H_0 , с эмпирической функцией распределения F_n , используются в качестве проверочных статистик различные меры "расстояния" между экспериментальной и гипотетической функциями распределения. Характерной особенностью непараметрических критериев является выбор в качестве указанной меры функционалов от F и F_n , распределения которых от F не зависят. Эти функционалы удобно строить на основе так называемого равномерного эмпирического процесса^{/1/}:

$$v_n(t) = \sqrt{n}(F_n(t) - t), \quad 0 \leq t \leq 1. \quad (1)$$

В основе использования функционалов от эмпирического процесса лежит слабая сходимость распределений этих функционалов к предельным распределениям^{/1, 2/}, являющаяся следствием сходимости по распределению процесса $v_n(t)$ к процессу $v(t)$, называемому броуновским мостом (см., например,^{/3/}).

К широкоизвестным статистикам такого типа относятся, например, статистика Колмогорова $D_n^{/4/}$, представляющая собой супремум от модуля эмпирического процесса

$$D_n = \sqrt{n} \sup_{0 \leq t \leq 1} |F(t) - t|,$$

статистика Смирнова — Крамера — Мизеса (СКМ) (см., например,^{/5/}), представляемая в виде интеграла от квадрата эмпирического процесса

$$w_n^2 = \int_0^1 [v_n(t)]^2 dt = n \int_0^1 [F_n(t) - t]^2 dt,$$

а также их модификации, например статистики D_n^- и D_n^+ , предложенные Смирновым^{/6, 7/}, и статистика Андерсона — Дарлингa^{/8/}, связанная с выбором весовой функции под знаком интеграла в выражении для w_n^2 . Эти статистики хорошо изучены, на их основе построены критерии

согласия (см., например, /9, 10/). К этому же классу можно отнести статистику, предложенную в работе /11/*:

$$\bar{F}_n = \int_0^1 F_n(t) dt = 1 - \frac{1}{n} \sum_{i=1}^n t_i,$$

поскольку она также может быть представлена через интеграл от эмпирического процесса:

$$w_n^1 = \int_0^1 v_n(t) dt = \sqrt{n} \int_0^1 [F_n(t) - t] dt = \sqrt{n} [F_n - \frac{1}{2}].$$

На практике применение того или иного непараметрического критерия зависит от исследуемой задачи, от особенностей конкретной статистики. Соотношение между функциями мощности для критериев, основанных на различных функционалах от $v_n(t)$, сильно зависит от параметров альтернативной гипотезы и установленного уровня значимости (см., например, /12/). Это обуславливает целесообразность выбора критерия согласия, наиболее подходящего для исследуемой задачи, стимулирует поиск и изучение новых критериев.

В настоящей работе предложена статистика, которая может быть представлена в виде интеграла от третьей степени эмпирического процесса $w_n^3 = \int_0^1 [v_n(t)]^3 dt$, и исследованы ее основные характеристики.

1. ОПРЕДЕЛЕНИЕ И АЛГЕБРАИЧЕСКИЙ ВИД СТАТИСТИКИ w_n^3

Рассмотрим интеграл

$$w_n^3 = n^{3/2} \int_{-\infty}^{\infty} [S_n(x) - P(x)]^3 dP(x), \quad (2)$$

где $P(x)$ — непрерывная функция распределения случайной величины x , n — объем выборки, а $S_n(x)$ — эмпирическая функция распределения величины x :

$$S_n(x) = 0 \quad \text{при } x < x_1,$$

$$S_n(x) = \frac{k}{n} \quad \text{при } x_k \leq x < x_{k+1} \quad (k = 1, 2, \dots, n-1), \quad (3)$$

* В более ранней работе /12/ упоминается статистика $\bar{U}_n = \frac{1}{n} \sum_{i=1}^n t_i$, предложенная Мозесом.

$$S_n(x) = 1 \quad \text{при } x \geq x_n,$$

здесь $x_1 < x_2 < x_3 < \dots < x_n$ — вариационный ряд, n — объем выборки.

Положим $t = P(x)$ ($t_i = P(x_i)$, $i = 1, 2, \dots, n$). Легко видеть, что величины t_1, t_2, \dots, t_n представляют собой вариационный ряд по выборке из равномерного на $[0, 1]$ распределения. Обозначим через $F_n(t)$ эмпирическую функцию распределения, вычисленную от t_1, t_2, \dots, t_n . Очевидно, что значения эмпирических функций $F_n(t)$ и $S_n(x)$ совпадают при $t = P(x)$. Тогда выражение (2) можно представить в виде

$$w_n^3 = n^{3/2} \int_0^1 [F_n(t) - t]^3 dt$$

или

$$w_n^3 = \int_0^1 [v_n(t)]^3 dt,$$

где $v_n(t) = \sqrt{n}(F_n(t) - t)$ — эмпирический процесс (1).

Таким образом, статистика w_n^3 может быть представлена в виде интеграла от третьей степени эмпирического процесса, что объясняет выбор коэффициента $n^{3/2}$ в формуле (2). Из возможности представления статистики w_n^3 через величины, не зависящие от $P(x)$, следует независимость распределения этой статистики от $P(x)$.

Для практических целей удобно использовать алгебраическое представление w_n^3 , вывод которого дан в приложении:

$$w_n^3 = -\frac{\sqrt{n}}{8} \sum_{i=1}^n [2P(x_i) - \frac{2i-1}{n}] \{ [2P(x_i) - \frac{2i-1}{n}]^2 + \frac{1}{n^2} \}.$$

Из интегрального выражения (2) для w_n^3 легко видеть, что для выборки объема n область изменения величины w_n^3 определяется интервалом $(-\frac{1}{4}n^{3/2}, \frac{1}{4}n^{3/2})$.

2. МОМЕНТЫ РАСПРЕДЕЛЕНИЯ: СРЕДНЕЕ ЗНАЧЕНИЕ И ДИСПЕРСИЯ

Среднее значение w_n^3 определяется из соотношения

$$M(w_n^3) = M \{ n^{3/2} \int_{-\infty}^{\infty} [S_n(x) - P(x)]^3 dP(x) \} = n^{3/2} \int_{-\infty}^{\infty} M \{ [S_n(x) - P(x)]^3 \} dP(x).$$

Для вычисления последнего интеграла введем индикатор

$$\xi_i = \begin{cases} 1 & \text{при } x_i < x, \\ 0 & \text{при } x_i > x, \end{cases}$$

где x_i — элемент случайной выборки x_1, x_2, \dots, x_n . Очевидно, что величины ξ_i, ξ_j независимы при $i \neq j$ ($i, j = 1, 2, \dots, n$), $(\xi_i)^3 = (\xi_j)^2 = \xi_i$,

а $M\xi_i = P(x)$ и $S_n(x) = \frac{1}{n} \sum_{i=1}^n \xi_i$. Тогда можно записать:

$$M\{[S_n(x) - P(x)]^3\} = \frac{1}{n^3} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n M\{(\xi_i - M\xi_i)(\xi_j - M\xi_j)(\xi_k - M\xi_k)\}.$$

Так как $M(\xi_i - M\xi_i) = 0$, то в последней сумме не равными нулю окажутся только члены, содержащие $M(\xi_i - M\xi_i)^3$, поэтому

$$M\{[S_n(x) - P(x)]^3\} = \frac{1}{n^3} \sum_{i=1}^n M(\xi_i - M\xi_i)^3 = \frac{1}{n^2} [P(x) + 2P^3(x) - 3P^2(x)].$$

После замены $P = P(x)$ для $M(w_n^3)$ получаем

$$M(w_n^3) = \frac{1}{\sqrt{n}} \int_0^1 (P + 2P^3 - 3P^2) dP = 0. \quad (4)$$

Перейдем теперь к вычислению $D(w_n^3)$:

$$D(w_n^3) = M\{(w_n^3)^2\} - [M(w_n^3)]^2.$$

Согласно (4) $M(w_n^3) = 0$, поэтому $D(w_n^3) = M\{(w_n^3)^2\}$. Таким образом, для определения дисперсии необходимо вычислить интеграл:

$$\begin{aligned} M\{(w_n^3)^2\} &= n^3 M\left\{ \int_{-\infty}^{\infty} [S_n(x^I) - P(x^I)]^3 [S_n(x^{II}) - P(x^{II})]^3 dP(x^I) dP(x^{II}) \right\} = \\ &= 2n^3 M\left\{ \int_{x^{II} > x^I} [S_n(x^I) - P(x^I)]^3 [S_n(x^{II}) - P(x^{II})]^3 dP(x^I) dP(x^{II}) \right\} = \\ &= 2n^3 \int_{x^{II} > x^I} M\{[S_n(x^I) - P(x^I)]^3 [S_n(x^{II}) - P(x^{II})]^3\} dP(x^I) dP(x^{II}). \end{aligned} \quad (5)$$

Введем индикаторы

$$\xi_i^I = \begin{cases} 1 & \text{при } x_i < x^I, \\ 0 & \text{при } x_i > x^I, \end{cases} \quad \xi_j^{II} = \begin{cases} 1 & \text{при } x_j < x^{II}, \\ 0 & \text{при } x_j > x^{II}, \end{cases} \quad (6)$$

где x_i, x_j — элементы случайной выборки x_1, x_2, \dots, x_n ; $i, j = 1, 2, \dots, n$. Подынтегральное выражение в (5) можно записать в виде

$$\begin{aligned} M\{[S_n(x^I) - P(x^I)]^3 [S_n(x^{II}) - P(x^{II})]^3\} &= \frac{1}{n^6} \sum_{i,j,k,\ell,s,t} M\{(\xi_i^I - M\xi_i^I)(\xi_j^I - M\xi_j^I) \times \\ &\times (\xi_k^I - M\xi_k^I)(\xi_\ell^{II} - M\xi_\ell^{II})(\xi_s^{II} - M\xi_s^{II})(\xi_t^{II} - M\xi_t^{II})\}, \end{aligned}$$

где суммирование ведется по шести индексам. Легко проверить, что в последней сумме не равны нулю только члены следующего вида (для удобства они обозначены через I_m , $m = 1, 2, \dots, 8$):

$$\begin{aligned} I_1 &= M\{(\xi_i^I - M\xi_i^I)^3\} M\{(\xi_j^{II} - M\xi_j^{II})^3\}, \quad i \neq j; \\ I_2 &= M\{(\xi_i^I - M\xi_i^I)^3 (\xi_i^{II} - M\xi_i^{II})^3\}; \\ I_3 &= M\{(\xi_i^I - M\xi_i^I)^2\} M\{(\xi_j^I - M\xi_j^I)(\xi_j^{II} - M\xi_j^{II})\} M\{(\xi_\ell^{II} - M\xi_\ell^{II})^2\}, \quad i \neq j, i \neq \ell, j \neq \ell; \\ I_4 &= M\{(\xi_i^I - M\xi_i^I)^2 (\xi_i^{II} - M\xi_i^{II})^2\} M\{(\xi_j^I - M\xi_j^I)(\xi_j^{II} - M\xi_j^{II})\}, \quad i \neq j; \\ I_5 &= M\{(\xi_i^I - M\xi_i^I)(\xi_i^{II} - M\xi_i^{II})\} M\{(\xi_j^I - M\xi_j^I)(\xi_j^{II} - M\xi_j^{II})\} \times \\ &\times M\{(\xi_k^I - M\xi_k^I)(\xi_k^{II} - M\xi_k^{II})\}, \quad i \neq j, i \neq k, j \neq k; \\ I_6 &= M\{(\xi_i^I - M\xi_i^I)^2\} M\{(\xi_j^I - M\xi_j^I)(\xi_j^{II} - M\xi_j^{II})^3\}, \quad i \neq j; \\ I_7 &= M\{(\xi_i^{II} - M\xi_i^{II})^2\} M\{(\xi_j^{II} - M\xi_j^{II})(\xi_j^I - M\xi_j^I)^3\}, \quad i \neq j; \\ I_8 &= M\{(\xi_i^I - M\xi_i^I)^2 (\xi_i^{II} - M\xi_i^{II})\} M\{(\xi_j^I - M\xi_j^I)(\xi_j^{II} - M\xi_j^{II})^2\}, \quad i \neq j. \end{aligned} \quad (7)$$

В результате получаем

$$\begin{aligned} M\{[S_n(x^I) - P(x^I)]^3 [S_n(x^{II}) - P(x^{II})]^3\} &= \frac{1}{n^6} (\sum_{i \neq j} I_1 + \sum_i I_2 + 9 \sum_{i \neq j, i \neq \ell, j \neq \ell} \sum_{j \neq \ell} I_3 + \\ &+ 9 \sum_{i \neq j} \sum_{i \neq j, i \neq \ell, j \neq \ell} \sum_{i \neq j} I_4 + 6 \sum_{i \neq j, i \neq \ell, j \neq \ell} \sum_{i \neq j} I_5 + 3 \sum_{i \neq j} \sum_{i \neq j} I_6 + 3 \sum_{i \neq j} \sum_{i \neq j} I_7 + 9 \sum_{i \neq j} \sum_{i \neq j} I_8) = \frac{1}{n^6} [n(n-1) I_1 + \\ &+ n I_2 + 9n(n-1)(n-2) I_3 + 9n(n-1) I_4 + 6n(n-1)(n-2) I_5 + \end{aligned} \quad (8)$$

$$+ 3n(n-1)I_6 + 3n(n-1)I_7 + 9n(n-1)I_8].$$

Из выражений для индикаторов ξ_1^I и ξ_1^{II} (см. (6)) следует, что $M\xi_1^{II} = P(x^{II})$, $M\xi_1^I = P(x^I)$, $(\xi_1^I)^3 = (\xi_1^I)^2 = \xi_1^I$, $(\xi_1^{II})^3 = (\xi_1^{II})^2 = \xi_1^{II}$ и для $x^{II} > x^I$ $\xi_j^I \cdot \xi_j^{II} = \xi_j^I$. Используя эти свойства, выражения (7) можно представить в виде

$$\begin{aligned} I_1 &= (P_1 + 2P_1^3 - 3P_1^2)(P_2 + 2P_2^3 - 3P_2^2); \\ I_2 &= P_1 + 3P_1P_2^2 - 3P_1P_2 - P_1P_2^3 + 3P_1^3 + 12P_1^3P_2^2 - 10P_1^3P_2 - 5P_1^3P_2^3 - 3P_1^2 - 9P_1^2P_2^2 + \\ &\quad + 9P_1^2P_2 + 3P_1^2P_2^3; \\ I_3 &= P_1^2(1 - P_2)^2P_2(1 - P_1); \\ I_4 &= P_1^2 - 3P_1^2P_2 - 2P_1^3 + 7P_1^3P_2 + 3P_1^2P_2^2 - 3P_1^3P_2^2 - 5P_1^3P_2^3 - P_1^2P_2^3 + 3P_1^3P_2^3; \\ I_5 &= P_1^3(1 - P_2)^3; \\ I_6 &= P_1^2 - 3P_1^2P_2^3 - 4P_1^2P_2 + 6P_1^2P_2^2 - P_1^3 + 3P_1^3P_2^3 + 4P_1^3P_2 - 6P_1^3P_2^2; \\ I_7 &= P_1P_2 + 3P_1^3P_2 - 3P_1^2P_2 - 2P_1P_2^2 - 6P_1^3P_2^2 + 6P_1^2P_2^2 + P_1P_2^3 + 3P_1^3P_2^3 - 3P_1^2P_2^3; \\ I_8 &= (P_1 - 2P_1^2 - P_1P_2 + 2P_1^2P_2)(P_1 - 3P_1P_2 + 2P_1^2P_2^2), \end{aligned} \quad (9)$$

где $P_1 = P(x^I)$, а $P_2 = P(x^{II})$.

Используя (8), выражение для дисперсии (5) можно записать в виде

$$\begin{aligned} D(w_n^3) &= 2n^3 \int_0^{P_2} dP_2 \int_0^{P_1} dP_1 \left\{ \frac{1}{n^6} [n(n-1)I_1 + nI_2 + 9n(n-1)(n-2)I_3 + \right. \\ &\quad + 9n(n-1)I_4 + 6n(n-1)(n-2)I_5 + 3n(n-1)I_6 + 3n(n-1)I_7 + \\ &\quad \left. + 9n(n-1)I_8] \right\}, \end{aligned}$$

где I_m ($m = 1, 2, \dots, 8$) представлены формулами (9).

После преобразований получаем

$$D(w_n^3) = \frac{40 - 96n + 79n^2}{1680n^2}.$$

Отсюда следует, что в асимптотическом пределе $n \rightarrow \infty$ $D(w^3) = 79/1680$. Эта величина представляет собой дисперсию асимптотического распределения статистики w^3 , доказательство существования которого полностью аналогично доказательству существования асимптотического распределения статистики w^2 , представленному в [9].

3. ПРОЦЕНТНЫЕ ТОЧКИ РАСПРЕДЕЛЕНИЯ w_n^3 ДЛЯ МАЛЫХ ВЫБОРОК

Размер области изменения величины w_n^3 сильно зависит от объема выборки n . Так, например, для $n = 1$ она определяется интервалом $(-0,25, 0,25)$, а для $n = 9$ — интервалом $(-6,75, 6,75)$ (см. раздел 1), но численные расчеты для $n = 1, 2, \dots, 9$ показывают, что вероятность случайной величине w_n^3 выйти за пределы $(-1,65, 1,65)$ не превышает 0,1%. Поэтому отрезок $[-1,65, 1,65]$ был выбран для вычисления функций распределения $\Phi_n(Z)$ для всех рассмотренных $n = 1, 2, \dots, 9$. Вычисления $\Phi_n(Z) = P\{w_n^3 < Z\}$ проводились в точках Z_i , $i = 1, 2, \dots, 101$, разбивающих указанный отрезок на 100 равных подынтервалов. Для расчетов, так же как в [14], использовался геометрический метод Монте-Карло. В каждом испытании генерировались n равномерно распределенных на $[0, 1]$ случайных чисел, которые располагались в виде вариационного ряда $x_1 < x_2 < \dots < x_n$ и использовались для вычисления величины $\xi \equiv w_n^3$ по формуле

$$\xi \equiv w_n^3 = -\frac{\sqrt{n}}{8} \sum_{i=1}^n \left(2x_i - \frac{2i-1}{n} \right) \left[\left(2x_i - \frac{2i-1}{n} \right)^2 + \frac{1}{n^2} \right]. \quad (10)$$

Значение функции распределения оценивалось как $\Phi_n(Z_j) = (N_{z_j}/N)$, где N — общее количество испытаний, а N_{z_j} — количество испытаний, удовлетворяющих условию $\xi < Z_j$. N выбиралось таким образом, чтобы обеспечить точность вычислений $\Phi_n(Z_j)$, равную единице третьего знака после запятой.

Затем для каждого n на основе полученных табличных зависимостей для функций распределения $\Phi_n(Z)$ путем кубической сплайн-интерполяции [15] были вычислены процентные точки распределений w_n^3 . Анализ таблиц процентных точек показал, что в пределах точности вычислений для всех рассмотренных Z справедливо равенство

$$\Phi_n(Z) = 1 - \Phi_n(-Z), \quad (11)$$

Таблица

Процентные точки Z_p случайной величины w_n^3 .

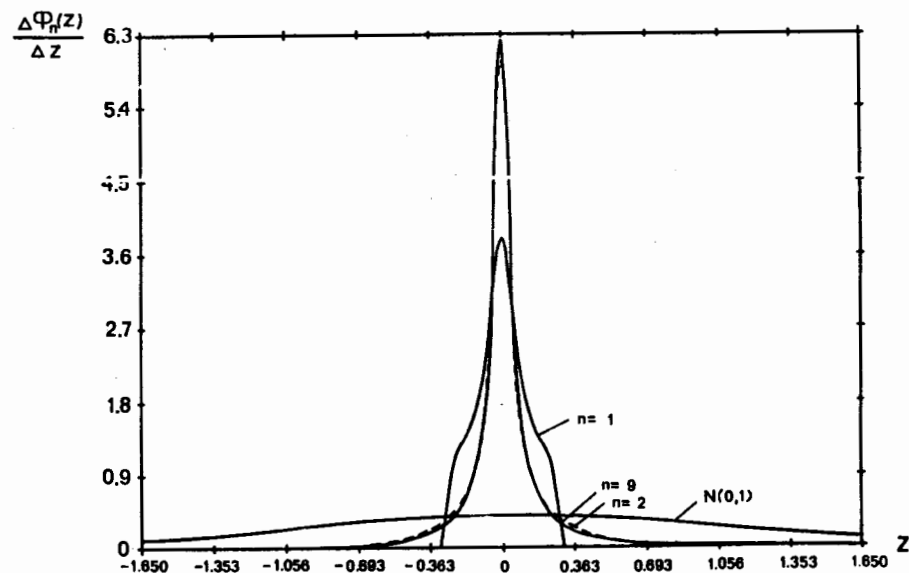
$$\Phi_n(Z_p) = P\{w_n^3 < Z_p\}$$

$\Phi_n \backslash n$	Процентные точки Z_p								
	1	2	3	4	5	6	7	8	9
.50	.000	.000	.000	.000	.000	.000	.000	.000	.000
.51	.003	.002	.001	.001	.001	.001	.001	.001	.001
.52	.005	.003	.003	.003	.003	.003	.003	.003	.003
.53	.008	.004	.004	.004	.004	.004	.004	.004	.004
.54	.010	.006	.005	.005	.005	.005	.005	.005	.005
.55	.013	.007	.007	.007	.007	.006	.006	.006	.006
.56	.015	.009	.008	.008	.008	.008	.008	.008	.008
.57	.018	.011	.009	.010	.010	.009	.009	.009	.009
.58	.021	.012	.011	.011	.011	.011	.011	.011	.011
.59	.023	.014	.013	.013	.013	.012	.012	.012	.012
.60	.026	.015	.014	.014	.014	.014	.014	.014	.014
.61	.029	.017	.016	.016	.016	.016	.016	.015	.015
.62	.032	.019	.018	.018	.018	.017	.017	.017	.017
.63	.035	.021	.020	.020	.020	.019	.019	.019	.019
.64	.038	.023	.022	.022	.021	.021	.021	.021	.021
.65	.041	.025	.024	.024	.024	.023	.023	.023	.023
.66	.044	.027	.026	.026	.026	.025	.025	.025	.025
.67	.047	.029	.028	.028	.028	.027	.027	.027	.027
.68	.051	.031	.031	.031	.030	.030	.030	.030	.029
.69	.054	.033	.034	.033	.033	.032	.032	.032	.032
.70	.058	.036	.036	.036	.035	.035	.035	.035	.035
.71	.062	.038	.039	.039	.039	.039	.039	.037	.037
.72	.066	.041	.043	.042	.041	.041	.041	.040	.040
.73	.070	.044	.046	.045	.044	.044	.044	.044	.043
.74	.074	.047	.050	.049	.048	.047	.047	.047	.047
.75	.078	.051	.053	.052	.051	.051	.051	.051	.050
.76	.083	.054	.058	.056	.055	.055	.055	.054	.054
.77	.087	.059	.062	.060	.059	.059	.059	.059	.058
.78	.092	.063	.067	.065	.064	.063	.063	.063	.063
.79	.097	.068	.071	.069	.068	.068	.068	.068	.068
.80	.102	.074	.077	.074	.074	.073	.073	.073	.073
.81	.107	.080	.082	.080	.079	.079	.079	.078	.078
.82	.113	.086	.088	.086	.085	.085	.085	.084	.084
.83	.119	.093	.095	.092	.092	.091	.091	.091	.091
.84	.124	.101	.102	.099	.099	.099	.098	.098	.098
.85	.131	.110	.110	.107	.107	.106	.106	.106	.106
.86	.137	.119	.118	.115	.115	.115	.114	.114	.114
.87	.143	.129	.127	.124	.125	.125	.124	.124	.124
.88	.150	.141	.137	.135	.135	.135	.135	.134	.134
.89	.157	.153	.148	.147	.147	.147	.146	.146	.146
.90	.164	.167	.161	.161	.161	.160	.160	.160	.160
.91	.171	.183	.175	.176	.176	.175	.175	.175	.175
.92	.179	.201	.191	.193	.193	.193	.193	.193	.193
.93	.19	.22	.21	.21	.21	.21	.21	.21	.21
.94	.20	.25	.23	.24	.24	.24	.24	.24	.24
.95	.20	.27	.26	.27	.27	.27	.27	.27	.27
.96	.21	.31	.30	.31	.31	.31	.31	.31	.31
.97	.22	.35	.35	.36	.36	.36	.37	.37	.37
.98	.23	.40	.43	.43	.44	.45	.45	.45	.45
.99	.24	.48	.55	.57	.58	.59	.59	.61	.61

которое является свойством симметричных относительно нуля распределений. Симметричность распределения w_n^3 подтверждается и результатами его моделирования (см. раздел 4)*. В таблице приведены 50 процентных точек для $Z > 0$, значения остальных процентных точек можно получить, используя (11). Точность приведенных в таблице значений с учетом возможных ошибок вычислений равна единице последнего знака.

4. МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ w_n^3

Моделирование распределения w_n^3 проводилось с целью извлечения дополнительной информации из анализа его формы. Для этого так же, как и в случае вычисления функций $\Phi_n(Z)$, генерировались n равномерно распределенных на $[0,1]$ случайных чисел x_i , $i = 1, 2, \dots, n$, которые располагались в виде вариационного ряда и использовались для вычис-



Распределение случайных величин w_n^3 для выборок $n = 1, 2$ и 9 .

* Строгое доказательство симметричности распределения w_n^3 не вызывает принципиальных затруднений, однако из-за его громоздкости авторы ограничились выводами, сделанными на основании численных расчетов.

ления w_n^3 по формуле (10). Полученные таким образом величины w_n^3 гистограммировались. Гистограммы нормировались на общее число событий с учетом шага гистограммирования.

На рисунке представлены эмпирические плотности распределений w_n^3 для $n = 1, 2$ и 9 , изображенные в виде кривых, аппроксимирующих соответствующие гистограммы; здесь же приведена плотность нормального распределения $N(0,1)$. Сравнение участков кривых, соответствующих положительным и отрицательным значениям w_n^3 , подтверждает предположение о симметричности формы распределения w_n^3 . Из рисунка также видно, что пик распределения w_n^3 находится выше пика кривой $N(0,1)$, поэтому эксцесс распределения w_n^3 положителен ($\epsilon > 0$).

ЗАКЛЮЧЕНИЕ

В работе предложена новая статистика, которая может быть представлена в виде интеграла от третьей степени эмпирического процесса $w_n^3 = \int_0^1 [v_n(t)]^3 dt$. Вычислены среднее значение и дисперсия распределения этой статистики, определены пределы изменения величины w_n^3 , получен алгебраический вид статистики, удобный для практических применений. В результате численного эксперимента установлено, что эксцесс распределения w_n^3 положителен, а форма распределения имеет симметричный вид. Вычислены таблицы процентных точек распределения случайной величины w_n^3 для малых объемов выборки: $n = 1, 2, \dots, 9$. Приведен вид эмпирической плотности распределения w_n^3 для выборок $n = 1, 2$ и 9 .

Авторы считают своим приятным долгом поблагодарить профессора Е.П.Жидкова за полезные обсуждения и поддержку, а также Н.И.Чернова за конструктивную критику и ценные замечания.

Приложение

АЛГЕБРАИЧЕСКИЙ ВИД СТАТИСТИКИ w_n^3

Разбивая область интегрирования в $w_n^3 = n^{3/2} \int_{-\infty}^{\infty} [S_n(x) - P(x)]^3 \times dP(x)$ на интервалы, как это сделано в^{13/} применительно к статистике $w_n^2, (-\infty, x_1), (x_1, x_2), \dots, (x_{n-1}, x_n), (x_n, \infty)$, можно записать

$$w_n^3 = -n^{3/2} \int_{-\infty}^{x_1} P^3(x) dP(x) + n^{3/2} \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} \left[\frac{i}{n} - P(x) \right]^3 dP(x) + n^{3/2} \int_{x_n}^{\infty} [1 - P(x)]^3 dP(x).$$

Учитывая, что $\lim_{x \rightarrow -\infty} P(x) = 0$, $\lim_{x \rightarrow \infty} P(x) = 1$, получаем

$$-n^{3/2} \int_{-\infty}^{x_1} P^3(x) dP(x) = -\frac{n^{3/2}}{4} P^4(x_1), \quad (1)$$

$$n^{3/2} \int_{x_n}^{\infty} [1 - P(x)]^3 dP(x) = \frac{n^{3/2}}{4} [1 - P(x_n)]^4, \quad (2)$$

$$n^{3/2} \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} \left[\frac{i}{n} - P(x) \right]^3 dP(x) = -\frac{n^{3/2}}{4} \left\{ \sum_{i=1}^n \left[\frac{i-1}{n} - P(x_i) \right]^4 - \sum_{i=1}^n \left[\frac{i}{n} - P(x_i) \right]^4 + [1 - P(x_n)]^4 - P^4(x_1) \right\}.$$

Нетрудно убедиться, что

$$\left[\frac{i}{n} - P(x_i) \right]^4 - \left[\frac{i-1}{n} - P(x_i) \right]^4 = \frac{1}{2n} \left[\frac{2i-1}{n} - 2P(x_i) \right] \left\{ [2P(x_i) - \frac{2i-1}{n}]^2 + \frac{1}{n^2} \right\},$$

поэтому

$$n^{3/2} \sum_{i=1}^{n-1} \int_{x_i}^{x_{i+1}} \left[\frac{i}{n} - P(x) \right]^3 dP(x) = -\frac{n^{3/2}}{4} \left\{ \frac{1}{2n} \sum_{i=1}^n \left[2P(x_i) - \frac{2i-1}{n} \right] \times \right. \\ \left. \times \left\{ [2P(x_i) - \frac{2i-1}{n}]^2 + \frac{1}{n^2} \right\} + [1 - P(x_n)]^4 - P^4(x_1) \right\}. \quad (3)$$

Суммируя правые части равенств (1), (2) и (3), получаем окончательное выражение для w_n^3 :

$$w_n^3 = -\frac{\sqrt{n}}{8} \sum_{i=1}^n \left[2P(x_i) - \frac{2i-1}{n} \right] \left\{ [2P(x_i) - \frac{2i-1}{n}]^2 + \frac{1}{n^2} \right\}.$$

ЛИТЕРАТУРА

1. Колмогоров А.Н. Теория вероятностей и математическая статистика. М.: Наука, 1986. Комментарий Хмаладзе Э.В. Эмпирические распределения.
2. Gaenssler P., Stute W. - The Annals of Probability, 1979, v.7, No2, p.193.

3. Боровков А.А. Теория вероятностей. М.: Наука, 1982.
4. Kolmogorov A.N. – G. Inst. Actuari., 1933, v.4, No1, p.83.
5. Статистические методы в экспериментальной физике. Пер. с англ. Под ред. А.А.Тяпкина. (Авт.: Идье В., Драйард Д., Джеймс Ф., Рус М., Садуле Б.). М.: Атомиздат, 1976, с.260.
6. Смирнов Н.В. Об отклонениях эмпирической кривой распределения. *Мат. сб.*, 1939, т.6(48), №1, с.3.
7. Смирнов Н.В. – УМН, 1944, вып.10, с.179.
8. Anderson T.W., Darling D.A. – *Ann. Math. Statist.*, 1952, v.23, p.193.
9. Durbin J. *Region Conf. in Appl. Math. 9th Issue. SIAM*, 1973.
10. Мартынов Г.В. Критерии омега-квадрат. М.: Наука, 1978.
11. Birnbaum Z.W., Tang V.K.T. – *Rev. Inst. Statist.*, 1964, 32, p.2.
12. Chapman D.G. – *Ann. Math. Statist.*, 1958, v.29, p.655.
13. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики. М.: Наука, 1965.
14. Зрелов П.В., Иванов В.В. Препринт ОИЯИ P10-86-547, Дубна, 1986.
15. Havie T. CERN Computer Centre Program Library E209.

Рукопись поступила в издательский отдел
12 мая 1988 года.

Зрелов П.В., Иванов В.В. P10-88-321
 Проверочная статистика $w_n^3 = n^{3/2} \int_{-\infty}^{\infty} [S_n(x) - P(x)]^3 dP(x)$
 в теории непараметрических критериев согласия

Предлагается новая статистика w_n^3 , которая может быть представлена в виде интеграла от третьей степени эмпирического процесса. Вычислены среднее значение и дисперсия распределения w_n^3 , получен алгебраический вид статистики, удобный для практических применений. На основании численных расчетов установлено, что эксцесс распределения w_n^3 положителен, а форма распределения имеет симметричный вид. Вычислены таблицы процентных точек распределения случайной величины w_n^3 для малых объемов выборки: $n = 1, 2, \dots, 9$. Приведен вид эмпирической плотности распределения w_n^3 для $n = 1, 2$ и 9 .

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1988

Перевод О.С.Виноградовой

Zrelov P.V., Ivanov V.V. P10-88-321
 Test Statistic $w_n^3 = n^{3/2} \int_{-\infty}^{\infty} [S_n(x) - P(x)]^3 dP(x)$
 in the Nonparametric Goodness-of-Fit Criteria Theory

The new w_n^3 -statistic presented as an integral of the third degree of the empirical process is proposed. The mean value and deviation of the w_n^3 -distribution are calculated. The algebraic form of the w_n^3 -statistic to be convenient for practical use is obtained. By numerical calculations it was found out that the excess of the w_n^3 -distribution is positive and the distribution has a symmetrical form. The percentage points of the w_n^3 -distribution for small sample sizes $n = 1, 2, \dots, 9$ are presented in tabulated form. The form of the empirical density of the w_n^3 -distribution for $n = 1, 2$ and 9 is given.

The investigation has been performed at the Laboratory of Computing Techniques and Automation, JINR.

Communication of the Joint Institute for Nuclear Research. Dubna 1988