сообщения
объединенного
института
ядерных
исследований
дубна

P11-92-398

П.В.Зрелов

ОБОБЩЕННЫЕ МОМЕНТНЫЕ ФУНКЦИИ ЭМПИРИЧЕСКОГО ПРОЦЕССА И ИНТЕГРАЛЬНЫЕ НЕПАРАМЕТРИЧЕСКИЕ СТАТИСТИКИ

1992

ВВЕДЕНИЕ

Понятие непараметрических критериев согласия тесно связано со свойством устойчивости. Этим свойством обладают критерии, размер критической области которых может быть рассчитан независимо от вида распределения, соответствующего проверяемой гипотезе. Это свойство обеспечивает большую степень универсальности таких критериев, а также простоту и удобство их использования.

Среди непараметрических критериев, предназначенных для обработки данных несгруппированных в гистограмму, особый интерес представляют критерии, основанные на сравнении функции распределения F_0, предсказываемой гипотезой H_0 , с эмпирической функцией распределения F_n при помощи проверочных статистик, представляющих собой различные меры "расстояния" между экспериментальной и гипотетической функциями распределения.* Как правило, рассматриваются критерии, принадлежащие двум основным типам - обобщенным критериям Колмогорова - Смирнова и Крамера - Мизеса :**

$$D_n(\psi) = \sqrt{n} \sup_{-\infty\le x\le\infty} (|F_n(x) - F_0(x)| \; \psi(F_0(x))), \tag{1}$$

$$W_n^2 = n \int_{-\infty}^{\infty} \psi(F_0(x))(F_n(x) - F_0(x))^2 dF_0(x), \tag{2}$$

где ψ - некоторая весовая функция.

С развитием теории случайных процессов, при исследовании свойств непараметрических критериев, статистики, лежащие в их основе, стали

---

*) Историю развития и исследования непараметрических статистик можно проследить по работам [6,9,10,11,12,14,15,16,20]. Этому вопросу посвящены монографии [2,7].

**) В литературе встречаются также и некоторые другие варианты названий для статистик типа D_n и W_n^2 (см., например, [2,6,7,23]). Кроме статистики D_n часто рассматриваются ее односторонние аналоги $D_n^{\pm}$.

записывать в виде функционалов от так называемого равномерного эмпирического процесса

$$y_n(t) = \sqrt{n}\ (F_n(t) - t), \tag{3}$$

где $F_n(t)$ - эмпирическая функция распределения по выборке объема n из равномерного на $[0,1]$ распределения : $t_1,\ t_2, \dots,\ t_n$.

С другой стороны, существуют критерии, которые не относятся ни к первому, ни ко второму типу. Прежде всего это критерии, основанные на статистике Мозеса $\bar{U}$ и предложенной в [24] статистике $\bar{F}_n$:

$$\bar{U} = 1/n \sum_{i=1}^{n} F(x_i), \qquad \bar{F}_n = 1 - 1/n \sum_{i=1}^{n} F(x_i).$$

Эти статистики могут быть представлены в виде интеграла от разности $[F_n(x) - F_0(x)]$ следующим образом

$$\sqrt{n} \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]dF_0(x) = \sqrt{n}\ [\bar{F}_n - 1/2],$$

$$\sqrt{n} \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]dF_0(x) = \sqrt{n}\ [1/2 - \bar{U}\ ].$$

Кроме того, в работах [4, 5] предложена и исследована статистика ω_n^3

$$\omega_n^3 = n^{3/2} \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^3\ dF_0(x),$$

также выходящая за рамки статистик типа омега - квадрат (W_n^2).

В настоящей работе показано, что многие из существующих непараметрических статистик типа (2), такие как ω_n^2 , A_n^2 , U_n^2, а также статистики $\bar{U}$ и ω_n^3 могут быть сведены под формальным определением обобщенной интегральной статистики

$$\omega_n(\varphi) = \int_0^1 \varphi\ [t, y_n(t)]\ dt, \tag{4}$$

где $y_n(t)$ - эмпирический процесс (3), а φ - некоторая функция этого процесса, интегрируемая по параметрам, представляемым элементами эмпирической выборки $t_1,\ t_2, \dots,\ t_n$. Кроме того, функция φ не должна выводить $\omega_n(\varphi)$ из класса функционалов, определенных на множестве функций $D_{[0,1]}$ и непрерывных в метрике Скорохода $d(x,y)^{*)}$.

В работе исследуются корреляционные свойства и моменты распределений статистик типа (4). Эти исследования основываются на изучении различных моментных функций эмпирического процесса. Так, в работе показана связь обобщенных моментных функций эмпирического процесса, обозначаемых в дальнейшем J_q:

$$J_q(t^{(I)}, t^{(II)}, \dots, t^{(q)}\ ; n, \varphi_1, \varphi_2, \dots, \varphi_q) =$$

$$= E\left\{ \varphi_1[t^I, y_n(t^I)]\varphi_2[t^{II}, y_n(t^{II})]\dots\ \varphi_q[t^{(q)}, y_n(t^{(q)})] \right\}, \tag{5}$$

где $\varphi_1,\ \varphi_2, \dots,\ \varphi_q$ - функции, требования к которым определены выше, с моментами и ковариациями статистик (4).

Функция (5) более подробно исследована в двух важных случаях :

$\varphi_i[t, y_n(t)] = [y_n(t)]^{k_i}$, $i = 1,2,\dots,\ q$ и $\varphi_i[t, y_n(t)] = |y_n(t)|^{k_i}$, $i = 1,2,\dots,q$. На этом основании исследованы статистики, представимые в виде

$$\omega(n,\ k) = \int_0^1 y_n^k(t)\ dt \tag{6}$$

и

$$\omega_A(n,\ k) = \int_0^1 |\ y_n(t)\ |^k dt, \tag{7}$$

где $k = 1,2,\dots$.

1. ЭМПИРИЧЕСКИЙ ПРОЦЕСС И ЕГО МОМЕНТНЫЕ ФУНКЦИИ

Эмпирический процесс имеет вид (см., например, [1]) :

$$y_n(t) = n^{1/2}\ (\ F_n(t) - t\ ), \tag{1.1}$$

где $0 \le t \le 1$, n - объем экспериментальной выборки, $F_n(t)$ - эмпирическая

---

* $D_{[0,1]}$ - совокупность вещественных функций на отрезке $[0,1]$, непрерывных справа и имеющих левосторонние пределы, $d(x,y) = inf_\lambda[sup_t|x(t) - y(\lambda(t))| + sup_t|t - \lambda(t)|]$, где $x(t),\ y(t) \in D_{[0,1]}$, $\lambda \in \Lambda$, а Λ - совокупность всех строго возрастающих непрерывных на $[0,1]$ функций, таких что $\lambda(0) = 0$, $\lambda(1) = 1$ [7]. В основе использования таких функционалов от эмпирического процесса при построении новых критериев лежит их сходимость по распределению к так называемому броуновскому мосту $(B^0(t))$: $g\ [y_n(t)] \longrightarrow g\ [B^0(t)]$ (см., например, [1], [13]).

функция распределения случайной величины t по выборке объема n : t_1, $t_2,\dots, t_n$ из равномерного на $[0,1]$ распределения . Эта функция может быть записана через индикатор $\xi_i(t)$:

$$\xi_i(t) = \begin{cases} 1, & t_i < t, \\ 0, & t_i > t, \quad i = 1,\ 2,\dots\ n \end{cases} \tag{1.2}$$

следующим образом

$$F_n(t) = 1/n \sum_{i=1}^{n} \xi_i(t) .$$

И, следовательно, для эмпирического процесса справедливо равенство

$$y_n(t) = 1/n^{1/2} \left\{ \sum_{i=1}^{n} (\xi_i(t) - t) \right\} . \tag{1.3}$$

Для моментной функции* эмпирического процесса $y_n(t)$ порядка $p' = k_1 + k_2 + \dots + k_q$ введем обозначение

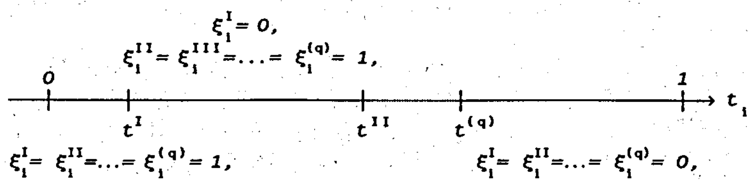
$$J_q(t^{I},t^{II},\dots,t^{(q)};\ n,k_1,k_2,\dots,k_q) = E\left\{ y_n^{k_1}(t^{I}) y_n^{k_2}(t^{II}) \dots y_n^{k_q}(t^{(q)}) \right\} =$$

$$= \int_0^1 \int_0^1 \dots \int_0^1 y_n^{k_1}(t^{I}) y_n^{k_2}(t^{II}) \dots y_n^{k_q}(t^{(q)}) dt_1 dt_2 \dots dt_n .$$

Для вычисления этой функции определим q индикаторов в виде

$$\xi_i(t^{(j)}) = \begin{cases} 1, & t_i < t^{(j)}, \\ 0, & t_i > t^{(j)};\ j = 1,2,\dots,q , \\ & i = 1,2,\dots,n . \end{cases}$$

Пусть $t^{I} < t^{II} < \dots < t^{(q)}$, тогда соотношение между индикаторами $\xi_i(t^{j})$, $j = 1,2,\dots,q$ для выбранного i можно установить при помощи следующей диаграммы



что можно записать в виде системы

*) Определение понятия моментной функции см., в частности, в [18].

$$\begin{cases} \xi_i^{I} = \xi_i^{II} = \dots = \xi_i^{(q)} = 1 & \text{при}\quad t_i \in (0,\ t^{I}), \\ \xi_i^{I} = 0;\ \xi_i^{II} = \xi_i^{III} = \dots = \xi_i^{(q)} = 1 & \text{при}\quad t_i \in (t^{I},\ t^{II}), \\ \xi_i^{I} = \xi_i^{II} = 0;\ \xi_i^{III} = \xi_i^{IV} = \dots = \xi_i^{(q)} = 1 & \text{при}\quad t_i \in (t^{II},\ t^{III}), \\ \dots & \\ \xi_i^{I} = \xi_i^{II} = \dots = \xi_i^{(q)} = 0 & \text{при}\quad t_i \in (t^{(q)},\ 1). \end{cases} \tag{1.4}$$

Разбивая область интегрирования для t_i , $i = 1,\ 2,\dots,\ n$ на $q+1$ частей $(0,\ t^{I})$, $(t^{I},\ t^{II}),\dots,\ (t^{(q)},\ 1)$ и используя (1.3), можно записать

$$J_q(t^{I},t^{II},\dots,t^{(q)};n,k_1,k_2,\dots,k_q) \overset{t^{I}<t^{II}<\dots<t^{(q)}}{=\!=\!=}$$

$$= 1/n^{(k_1\ k_2+\dots+k_q)/2} \left\{ \int_0^{t^{I}} + \int_{t^{I}}^{t^{II}} + \dots + \int_{t^{(q)}}^{1} \right\} \dots \left\{ \int_0^{t^{I}} + \int_{t^{I}}^{t^{II}} + \dots + \int_{t^{(q)}}^{1} \right\} \times$$

$$\times dt_1 dt_2 \dots dt_n \left\{ \sum_{i=1}^{n} (\xi_i^{I} - t^{I}) \right\}^{k_1} \left\{ \sum_{j=1}^{n} (\xi_j^{II} - t^{II}) \right\}^{k_2} \dots \left\{ \sum_{m=1}^{n} (\xi_m^{(q)} - t^{(q)}) \right\}^{k_q} .$$

Выпишем общий член I_q величины J_q , используя свойства (1.4):

$$I_q(t^{I},t^{II},\dots,t^{(q)};n,k_1,k_2,\dots,k_q) = 1/n^{(k_1\ k_2+\dots+k_q)/2} \times$$

$$\times \underbrace{\int_0^{t^{I}} \dots \int_0^{t^{I}}}_{s_1} \underbrace{\int_{t^{I}}^{t^{II}} \dots \int_{t^{I}}^{t^{II}}}_{s_2} \dots \underbrace{\int_{t^{(q)}}^{1} \dots \int_{t^{(q)}}^{1}}_{n-(s_1+s_2+\dots+s_q)} dt_1 dt_2 \dots dt_n \left\{ \sum_{i=1}^{n} (\xi_i^{I} - t^{I}) \right\}^{k_1} \dots \times$$

$$\times \left\{ \sum_{m=1}^{n} (\xi_m^{(q)} - t^{(q)}) \right\}^{k_q} = 1/n^{(k_1\ k_2+\dots+k_q)/2} (t^{I})^{s_1} (t^{II} - t^{I})^{s_2} \dots \times$$

$$\times (t^{(q)} - t^{(q-1)})^{s_q} (1 - t^{(q)})^{n-\sum_{i=1}^{q} s_i} \left\{ s_1(1 - t^{I}) + (n - s_1)(-t^{I}) \right\}^{k_1} \times$$

$$\times \left\{ (s_1 + s_2)(1 - t^{II}) + [n - (s_1 + s_2)](-t^{II}) \right\}^{k_2} \dots \left\{ (\sum_{i=1}^{q} s_i)(1 - t^{(q)}) + \right.$$

$$\left. + (n - \sum_{i=1}^{q} s_i)(-t^{(q)}) \right\}^{k_q} . \tag{1.5}$$

С учетом перестановок интегралов связь между J_q и I_q определяется равенством

$$J_q(t^{I},t^{II},\ldots,t^{(q)};n,k_1,k_2,\ldots,k_q) = \sum_{s_1=0}^{n}\ \sum_{s_2=0}^{n-s_1}\cdots\sum_{s_q=0}^{n-(s_1+s_2+\ldots+s_{q-1})}\times$$

$$\times\, C_n^{s_1}\, C_{n-s_1}^{s_2}\ \ldots\ C_{n-(s_1+s_2+\ldots+s_{q-1})}^{s_q}\ I_q(t^{I},t^{II},\ldots,t^{(q)};n,k_1,k_2,\ldots,k_q),$$

где C_m^r - биномиальные коэффициенты. Учитывая (*1.5*), для J_q окончательно получаем

$$J_q(t^{I},t^{II},\ldots,t^{(q)};n,k_1,k_2,\ldots,k_q) \overset{t^{I}<t^{II}<\ldots<t^{(q)}}{=\!=\!=}$$

$$= 1/n^{(k_1\ k_2+\ldots+k_q)/2}\sum_{s_1=0}^{n}\ \sum_{s_2=0}^{n-s_1}\cdots\sum_{s_q=0}^{n-(s_1+s\ +\ldots+s_{q-1})} C_n^{s_1}\, C_{n-s_1}^{s_2}\ \ldots\times$$

$$\times\, C_{n-(s_1+s_2+\ \ldots+s_{q-1})}^{s_q}\ (t^{I})^{s_1}(t^{II}-t^{I})^{s_2}\ldots\ (1-t^{(q)})^{n-(s_1+s_2+\ldots s_q)}\times$$

$$\times\ (s_1-nt^{I})^{k_1}(s_1+s_2-nt^{II})^{k_2}\ldots\ (\sum_{i=1}^{q}s_i-\ t^{(q)})^{k_q}. \qquad (1.6)$$

Рассмотрим два частных случая (*1.6*).

а) Среднее значение k - й степени эмпирического процесса

$$J_1\ (t;\ n,k) = E\left\{\ y_n^k(t)\ \right\} = 1/n^{k/2}\sum_{s=0}^{n}C_n^s\ t^s\ (1-t)^{n-s}\ (s-tn)^k \qquad (1.7)$$

б) Корреляционная функция к - й степени эмпирического процесса

$$J_2(\ t,t';n,k\ ) = E\left\{\ y_n^k(t)\ y_n^k(t')\right\} \overset{t<t'}{=\!=\!=} 1/n^k\sum_{s_1=0}^{n}\ \sum_{s_2=0}^{n-s_1}\ C_n^{s_1}\ C_{n-s_1}^{s_2}\times$$

$$\times\, t^{s_1}(\ t'-t\ )^{s_2}(1-t')^{n-(s_1+s_2)}\ \ (\ s_1-tn\ )^k(\ s_1+s_2-t'n\ )^k. \qquad (1.8)$$

Введем теперь понятие обобщенной моментной функции эмпирического процесса при помощи следующего равенства

$$J_q(t^{I},t^{II},\ldots,t^{(q)};\ n,\varphi_1,\varphi_2,\ldots,\varphi_q)=E\Big\{\varphi_1(t^{I},y_n(t^{I}))\ldots\varphi_q(t^{(q)},y_n(t^{(q)}))\Big\}, \qquad (1.9)$$

где $\varphi_1,\ \varphi_2,\ldots,\ \varphi_q$ - некоторые функции процесса (*1.1*), интегрируемые по параметрам, представляемым элементами эмпирической выборки $t_1,\ t_2,\ldots,\ t_n$. В случае $\varphi_i(t,y_n(t)) = y_n^{k_i}(t),\ i = 1,\ 2,\ldots,\ q$ равенство (*1.9*) определяет моментную функцию эмпирического процесса порядка $p' = k_1 + k_2 + \ldots + k_q$ (*1.6*).

В силу указанной интегрируемости функций φ_i для моментной функции (*1.9*) путем обобщения (*1.6*) можно записать

$$J_q(t^{I},t^{II},\ldots,t^{(q)};\ n,\varphi_1,\varphi_2,\ldots,\varphi_q) \overset{t^{I}<t^{II}<\ldots<t^{(q)}}{=\!=\!=}$$

$$= \sum_{s_1=0}^{n}\ \sum_{s_2=0}^{n-s_1}\cdots\sum_{s_q=0}^{n-(s_1+s_2+\ldots s_{q-1})} C_n^{s_1}\ C_{n-s_1}^{s_2}\ldots C_{n-(s_1+s_2+\ldots s_{q-1})}^{s_q}\times$$

$$\times\ (t^{I})^{s_1}(t^{II}-t^{I})^{s_2}\ldots(1-t^{(q)})^{n-(s_1+s_2+\ldots+s_q)}\times$$

$$\times\ \varphi_1[t^{I},n^{-1/2}(s_1-nt^{I})]\ldots\varphi_q[t^{(q)},n^{-1/2}(\sum_{i=1}^{n-1}s_i-nt^{(q)})]. \qquad (1.10)$$

Обозначая функцию (*1.9*) для случая $\varphi_i(t,y_n(t)) = |y_n(t)|^{k_i},\ i = 1,\ 2,\ldots,q$, через J_q^A, имеем

$$J_q^A(t^{I},t^{II},\ldots,t^{(q)};\ n,k_1,k_2,\ldots,k_q) =$$

$$= E\left\{\ |\ y_n(t^{I})|^{k_1}|\ y_n(t^{II})|^{k_2}\ldots\ |\ y_n(t^{(q)})|^{k_q}\right\} \overset{t^{I}<t^{II}<\ldots<t^{(q)}}{=\!=\!=}$$

$$= 1/n^{(k_1+k_2+\ldots+k_q)/2}\sum_{s_1=0}^{n}\ \sum_{s_2=0}^{n-s_1}\cdots\sum_{s_q=0}^{n-(s_1+s_2+\ldots+s_{q-1})} C_n^{s_1}\ C_{n-s_1}^{s_2}\times\ldots$$

$$\times\ C_{n-(s_1+s_2+\ldots+s_{q-1})}^{s_q}\ (t^{I})^{s_1}(t^{II}-t^{I})^{s_2}\ldots\ (1-t^{(q)})^{n-(s_1+s_2+\ldots s_q)}\times$$

$$\times\ |\ s_1-nt^{I}|^{k_1}|\ s_1+s_2-nt^{II}|^{k_2}\ldots\ |\sum_{i=1}^{q}s_i-nt^{(q)}|^{k_q}\ . \qquad (1.11)$$

2. ОБОБЩЕННЫЕ СТАТИСТИКИ $\omega_n(\varphi)$

Обобщая известные статистики $\bar{U}$, U_n^2, ω_n^2, A_n^2 и ω_n^3 (см., например, [2], [3] и [4]), рассмотрим следующий функционал от эмпирического процесса:

$$\omega_n(\varphi) = \int_0^1 \varphi\ [t, y_n(t)]dt, \qquad (2.1)$$

где $\varphi(t, x)$ - некоторая функция, измеримая относительно произведения σ - алгебры борелевских множеств на $[0,1]$ и σ - алгебры борелевских множеств в фазовом пространстве X эмпирического процесса.

Если функция $E\ (\varphi[t^I, y_n(t^I)]\varphi[t^{II}, y_n(t^{II})]\ \ldots\ \varphi[t^{(q)}, y_n(t^{(q)})])$ от t_1, $t_2, \ldots, t_q$ является борелевской, можно записать*

$$M_q[\omega_n(\varphi)\ ] = M_q[\ \int_0^1 \varphi[t, y_n(t)]dt\ ] = E\left(\underbrace{\int_0^1 \ldots \int_0^1}_{q} dt^I dt^{II} \ldots dt^{(q)} \times\right.$$

$$\left.\times\ \varphi[t^I, y_n(t^I)] \ldots \varphi[t^{(q)}, y_n(t^{(q)})]\right) = \int_0^1 \ldots \int_0^1 dt^I dt^{II} \ldots dt^{(q)} \times$$

$$\times\ E\Big(\ \varphi[t^I, y_n(t^I)]\varphi[t^{II}, y_n(t^{II})]\ \ldots \varphi[t^{(q)}, y_n(t^{(q)})]\ \Big), \qquad (2.2)$$

где $M_q[\omega_n(\varphi)\ ]$ - нецентральный момент степени q статистики $\omega_n(\varphi)$.

Величина $E(\varphi[t^I, y_n(t^I)]\varphi[t^{II}, y_n(t^{II})] \ldots \varphi[t^{(q)}, y_n(t^{(q)})])$ представляет собой обобщенную моментную функцию эмпирического процесса J_q (1.9). Если φ отвечает дополнительному условию гл. 1, то формула (1.10) определяет J_q для случая $t^I < t^{II} < \ldots < t^{(q)}$ при $\varphi_1 = \varphi_2 = \ldots = \varphi_q = \varphi$ и легко может быть записана для любого другого соотношения между t^I, t^{II}, ..., $t^{(q)}$. Таким образом, (2.2) принимает вид суммы $q!$ кратных интегралов

$$M_q[\omega_n(\varphi)] = \int_0^1 dt^{(q)} \int_0^{t^{(q)}} dt^{(q-1)} \ldots \int_0^{t^{II}} dt^I\ J_q(t^I < t^{II} < \ldots < t^{(q)}; n, \varphi, \ldots, \varphi)\ +$$

*) См., например, В.С. Королюк и др. Справочник по теории вероятностей и математической статистики. М.:Наука, 1985, стр.212.

$$+ \int_0^1 dt^{(q)} \int_0^{t^{(q)}} dt^{(q-1)} \ldots \int_0^{t^{III}} dt^I \int_0^{t^I} dt^{II} J_q(t^{II} < t^I < t^{III} < t^{IV} < \ldots < t^{(q)}; n, \varphi, \ldots, \varphi)\ +$$

$$+\ \ldots\ + \int_0^1 dt^I \int_0^{t^I} dt^{II} \ldots \int_0^{t^{(q-1)}} dt^{(q)}\ J_q(t^{(q)} < t^{(q-1)} < \ldots < t^I; n, \varphi, \ldots, \varphi). \qquad (2.3)$$

Необходимое для вычисления ковариации и коэффициента корреляции среднее значение произведения $\omega_n(\varphi_1)$ и $\omega_n(\varphi_2)$ определяется аналогично (2.3) следующим образом :

$$E\ [\ \omega_n(\varphi_1)\omega_n(\varphi_2)\ ] = \int_0^1 dt^{II} \int_0^{t^{II}} dt^I\ J_2(\ t^I < t^{II};\ n, \varphi_1, \varphi_2)\ +$$

$$+ \int_0^1 dt^I \int_0^{t^I} dt^{II}\ J_2(\ t^{II} < t^I;\ n, \varphi_1, \varphi_2), \qquad (2.4)$$

где J_2 - обобщенные моментные функции (1.10) для $q = 2$ при соответствующих соотношениях между t^{II} и t^I.

Приведенные выше формулы позволяют исследовать моментные и корреляционные свойства статистик $\omega_n(\varphi)$ для широкого класса функций φ.

Перейдем теперь к исследованию некоторых типов интегральных статистик в рамках формального определения $\omega_n(\varphi)$.

Рассмотрим статистику

$$\omega(n,k) = n^{k/2} \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^k dF_0(x),$$

где $F_0(x)$, $F_n(x)$ - соответственно функция распределения нулевой гипотезы H_0 и эмпирическая функция распределения, построенная по выборке объема n из распределения с.в. x. Эта статистика может быть записана в виде интеграла от эмпирического процесса k - й степени

$$\omega(n,k) = \int_0^1 y_n^k(t)dt. \qquad (2.5)$$

Нецентральный момент степени q распределения $\omega(n,k)$ определяется формулой (2.3), в которой обобщенная функция J_q представляется выражением (1.6) при $k_1 = k_2 = \ldots = k_q = k$.

3. КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ И КОВАРИАЦИИ СТАТИСТИК $\omega(n,k)$

Запишем выражения для ковариации (C) и коэффициента корреляции (ρ) статистик $\omega(n,k_1)$ и $\omega(n,k_2)$:

$$C\,[\omega(n,k_1)\omega(n,k_2)] = E\,[\omega(n,k_1)\omega(n,k_2)] - E\,[\omega(n,k_1)]E\,[\omega(n,k_2)]$$

$$\rho = C\,[\omega(n,k_1)\omega(n,k_2)] \,/\, \sqrt{D\,[\omega(n,k_1)]\;D\,[\omega(n,k_2)]}, \tag{3.1}$$

где E - среднее значение, а дисперсия D статистики $\omega(n,k)$ определяется по формуле

$$D\,[\omega(n,k)] = E\,\{[\omega(n,k)]^2\} - E^2[\omega(n,k)]. \tag{3.2}$$

На основании изложенного в предыдущей главе получаем

$$E\,[\,\omega(n,k)\,] = \frac{1}{n^{k/2}} \sum_{s=0}^{n} C_n^s \int_0^1 t^s\,(1-t)^{n-s}\,(s-tn)^k\,dt. \tag{3.3}$$

Формулу (3.3) можно представить в алгебраической записи, преобразуя подынтегральные выражения с использованием формулы бинома Ньютона и интегрируя :

$$E\,[\omega(n,k)] = \frac{1}{n^{k/2}} \sum_{s=0}^{n} \sum_{m_1=0}^{n-s} \sum_{m_2=0}^{k} C_n^s\, C_{n-s}^{m_1}\, C_k^{m_2}\, \frac{(-1)^{m_1+m_2}\, n^{m_2}\, s^{k-m_2}}{m_1+m_2+s+1}, \tag{3.4}$$

где полагается $0^0 = 1$.

Аналогично среднее значение произведения $\omega(n,k_1)$ и $\omega(n,k_2)$ равно :

$$E\,[\,\omega(n,k_1)\omega(n,k_2)\,] = \int_0^1 dt^{II} \int_0^{t^{II}} dt^{I}\, J_2(\,t^I < t^{II};\, n,k_1,k_2) +$$

$$+ \int_0^1 dt^{I} \int_0^{t^{I}} dt^{II}\, J_2(\,t^{II} < t^{I};\, n,k_1,k_2), \tag{3.5}$$

где через $J_2(\,t^I < t^{II};\, n,k_1,k_2)$ и $J_2(\,t^{II} < t^{I};\, n,k_1,k_2)$ обозначены функции $J_2(\,t^I,\, t^{II};\, n,k_1,k_2)$ при соответствующих условиях.

В алгебраической записи

$$E\,[\omega(n,k_1)\omega(n,k_2)] = \frac{1}{n^{(k_1+k_2)/2}} \sum_{s_1=0}^{n} \sum_{s_2=0}^{n-s_1} \sum_{m_1=0}^{s_2} \sum_{m_2=0}^{n-s_1-s_2} \sum_{m_3=0}^{k_1} \sum_{m_4=0}^{k_2} C_n^{s_1} \times$$

$$\times\, C_{n-s_1}^{s_2}\, C_{s_2}^{m_1}\, C_{n-s_1-s_2}^{m_2}\, C_{k_1}^{m_3}\, C_{k_2}^{m_4}\, (-1)^{m_1+m_2+m_3+m_4}\, n^{m_3+m_4}\, \frac{1}{(s_1+s_2+m_2+m_3+m_4+2)} \times$$

$$\times \left(\frac{s_1^{k_1-m_3}\,(s_1+s_2)^{k_2-m_4}}{s_1+m_1+m_3+1} + \frac{s_1^{k_2-m_4}\,(s_1+s_2)^{k_1-m_3}}{s_1+m_1+m_4+1} \right), \tag{3.6}$$

где полагается $0^0 = 1$.

В случае $k_1 = k_2 = k$ формула (3.6) определяет второй момент $M_2[\omega(n,k)]$ статистики $\omega(n,k)$.

Алгебраический вид зависимости ковариаций и коэффициентов корреляции статистик $\omega(n,k_1)$ и $\omega(n,k_2)$, полученный на основе формул (3.1), (3.2), (3.4) и (3.6), использовался для реализации вычислений на фортране для $n = 1,2,3,4$ и $k_1,k_2 = 1(1)10$.

В таблицах 1 и 2 приведены результаты вычислений для случая $n = 4$.

Таблица 1. Коэффициенты корреляции статистик ω (n,k1) и ω (n,k2) при n = 4

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10
1	1	0	.8876	0	.6903	0	.5179	0	.3914	0
2		1	0	.9341	0	.8029	0	.6680	0	.5528
3			1	0	.9320	0	.8023	0	.6739	0
4				1	0	.9571	0	.8666	0	.7669
5					1	0	.9591	0	.8760	0
6						1	0	.9717	0	.9101
7							1	0	.9747	0
8								1	0	.9812
9									1	0
10										1

Таблица 2. Ковариации статистик ω (n,k1) и ω (n,k2) при n = 4

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10
1	.08333	0	.04583	0	.04241	0	.05356	0	.08348	0
2		.01806	0	.02019	0	.02737	0	.04425	0	.08202
3			.03199	0	.03548	0	.05141	0	.08906	0
4				.02587	0	.03905	0	.06871	0	.13620
5					.04530	0	.07312	0	.13770	0
6						.06435	0	.12150	0	.25490
7							.12830	0	.25800	0
8								.24300	0	.53400
9									.54580	0
10										1.2190

Диагональные элементы таблицы 2 представляют собой значения дисперсий распределений соответствующих статистик $\omega(n,k)$ при $n = 4$. Сравнение этих значений для $k = 2$ и $k = 3$ с вычисленными по известным формулам для статистик ω_n^2 и ω_n^3 [4]

$$D(\omega_n^2) = (4n - 3)/\ 180n\ ,\quad D(\omega_n^3) = (20 - 66n + 69n^2)/1680n^2 \tag{3.7}$$

показывает полное совпадение результатов.

Другой подход при вычислении ковариаций и коэффициентов корреляции состоит в использовании символьного интегрирования при реализации формул (3.3) и (3.5) (система аналитических вычислений *muMATH*). Были проведены вычисления для $k_1,\ k_2 = 1(1)5$ при нескольких значениях объема выборки n. Полученные значения полностью согласуются с результатами вычислений, проведенных на основе формул (3.4) и (3.6) (см. выше).

Особенностью такого подхода является тот факт, что получаемые значения, в отличие от значений ρ и C в таблицах 1 и 2, представленных в виде десятичных дробей, являются точными. Это обстоятельство может быть использовано для определениия аналитического вида зависимости от n для коэффициентов корреляции и ковариаций непараметрических статистик типа $\omega(n,k)$ для ряда конкретных значений k.

4. ОБОБЩЕНИЕ РЕЗУЛЬТАТОВ, ПОЛУЧЕННЫХ С ПОМОЩЬЮ СИСТЕМЫ muMATH

На основании анализа вида зависимостей от n средних значений и дисперсий статистик ω_n^2 и ω_n^3 делаем предположение, что дисперсия интегральной статистики $\omega(n,k)$ представляется в виде полинома степени $m = k - 1$ по $1/n$, а среднее значение - в виде полинома степени $m = k/2 - 1$ по $1/n$ для четных k и 0 - для нечетных.

Для поиска вида зависимости от n функции ковариации статистик $\omega(n,k_1)$ и $\omega(n,k_2)$ также предполагаем, что она представляется полиномом степени $k_1 \le m < k_2$ по $1/n$.

Обозначим через $\rho(n,k_1,k_2)$ и $C(n,k_1,k_2)$ соответственно коэффициент корреляции и ковариацию статистик $\omega(n,k_1)$ и $\omega(n,k_2)$.

Рассмотрим методику определения вида зависимостей $C(n,k_1,k_2)$ от n на примере статистик $\omega(n,1)$ и $\omega(n,3)$.

Положим $m = 2$ и составим систему линейных уравнений относительно коэффициентов искомого полинома $a + b/n + c/n^2$. Для этого вычислим значения ковариаци $C(n,1,3)$ при $n = 1,2$ и 3 : $C(1,3,1) = 1/30$, $C(2,3,1) = 1/24$, $C(3,3,1) = 2/45$. После чего запишем

$$\begin{cases} a + b + c - 1/30 = 0, \\ a + b/2 + c/4 - 1/24 = 0, \\ a + b/3 + c/9 - 2/45 = 0. \end{cases}$$

Система имеет единственное решение $a = 1/20$, $b = -1/60$, $c = 0$. Таким образом, искомая аналитическая зависимость представляется полиномом степени $m = 1$:

$$C\ [\omega(n,1)\omega(n,3)] = 1/20 - 1/60n = (\ 3n - 1\ )/\ 60n\ . \tag{4.1}$$

Значение этого полинома в точке $n = 4$ равно $11/240$, что совпадает со значением функции корреляции статистик ω_n^1 и ω_n^3 , вычисленным при помощи системы *muMATH* для $n = 4$.

Учитывая вид зависимости для $D(\omega_n^3)$ (см. гл.3), а также то, что $D(\omega_n^1) = 1/12$, на основании (4.1) получаем аналитическую зависимость от n коэффициента корреляции статистик ω_n^1 и ω_n^3 :

$$\rho\ [\omega(n,1)\omega(n,3)] = \frac{2/5\sqrt{35}\ (\ 3n - 1\ )}{\sqrt{20 - 66n + 69n^2}}\ . \tag{4.2}$$

Легко видеть, что в асимптотическом пределе ($n \longrightarrow \infty$) коэффициент корреляции (4.2) равен

$$\rho\ (\omega^1\omega^3) = 6\sqrt{35}\ /\ 5\sqrt{69} \cong 0.854675.$$

Рассмотрим теперь статистику ω_n^4 и ее связь с ω_n^2 :

$$E\ (\omega_n^4) = 1/10 - 1/\ 30n = (\ 3n - 1\ )/\ 30n,$$

$$D\ (\omega_n^4) = (\ 188n^3 - 382n^2 + 276n - 70\ )/\ 4200n^3,$$

$$C\ [\omega(n,2)\omega(n,4)] = (\ 36n^2 - 46n + 15\ )/\ 1260n^2. \tag{4.3}$$

Учитывая (4.3), а также вид зависимости $D(\omega_n^2)$ (см. гл.3), получаем для коэффициента корреляции статистик ω_n^2 и ω_n^4 :

$$\rho\ [\omega(n,2)\omega(n,4)] = \frac{2\sqrt{470}\ (\ 36n^2 - 46n + 15\ )}{63\sqrt{752n^4 - 2092n^3 + 2250n^2 - 1108n + 210}}\ . \tag{4.4}$$

Значение коэффициента (4.4) в асимптотическом пределе ($n \longrightarrow \infty$) равно

$$\rho\ (\ \omega^2\omega^4\ ) = 8\sqrt{235}\ /\ 7\sqrt{376} \cong 0.903508.$$

Кроме того, рассмотренным выше методом были вычислены средние значения ряда статистик :

$$E(\omega_n^6) = (9n^2 - 8n + 2)/(84n^2),$$

$$E(\omega_n^8) = (30n^3 - 50n^2 + 30n - 6)/(180n^3),$$

$$E(\omega_n^{10}) = (45n^4 - 120n^3 + 127n^2 - 60n + 10)/(132n^4),$$

$$E(\omega_n^{12}) = (4725n^5 - 18375n^4 + 29890n^3 - 24472n^2 + 9674n - 1382)/$$

$$/(5460n^5), \qquad (4.5)$$

для нечетных значений k средние значения равны нулю.

Перейдем теперь к рассмотрению некоторых модификаций статистик $\omega(n,k)$.

5. СТАТИСТИКА $\omega_A(n,k) = n^{k/2} \int_{-\infty}^{\infty} |F_n(x) - F_0(x)|^k dF_0(x)$

Запишем статистику $\omega_A(n,k)$ в виде интеграла от модуля k - й степени эмпирического процесса:

$$\omega_A(n,k) = \int_0^1 |y_n(t)|^k dt.$$

Моменты распределения статистики $\omega_A(n,k)$ определяются через интегралы соответствующей кратности от моментной функции J_q^A (1.11) аналогично статистике $\omega(n,k)$:

$$M_q[\omega_A(n,k)] = \int_0^1 dt^{(q)} \int_0^{t^{(q)}} dt^{(q-1)} \ldots \int_0^{t^{II}} dt^I J_q^A(t^I < t^{II} < \ldots < t^{(q)}; n,k,\ldots,k) +$$

$$+ \ldots + \int_0^1 dt^I \int_0^{t^I} dt^{II} \ldots \int_0^{t^{(q-1)}} dt^{(q)} J_q^A(t^{(q)} < t^{(q-1)} < \ldots < t^I; n,k,\ldots,k). \qquad (5.1)$$

Ковариация и коэффициент корреляции статистик $\omega_A(n,k_1)$ и $\omega_A(n,k_2)$ имеют вид :

$$C[\omega_A(n,k_1)\omega_A(n,k_2)] = E[\omega_A(n,k_1)\omega_A(n,k_2)] - E[\omega_A(n,k_1)]E[\omega_A(n,k_2)]$$

$$\rho = C[\omega_A(n,k_1)\omega_A(n,k_2)] / \sqrt{D[\omega_A(n,k_1)]\, D[\omega_A(n,k_2)]}, \qquad (5.2)$$

где D - обозначение дисперсии, а E - обозначение среднего значения.

Для вычисления моментов, ковариаций и коэффициентов корреляции была

составлена программа на аналитическом языке *muMATH*, реализующая указанные выше формулы. В таблице 3 приведены результаты вычислений по этой программе коэффициентов корреляции статистик $\omega_A(n,k_1)$ и $\omega_A(n,k_2)$ для объема выборки $n = 2$ и значений k_1, $k_2 = 1(1)5$.

Таблица 3. Коэффициенты корреляции статистик $\omega_A(n,k1)$ и $\omega_A(n,k2)$ при $n = 2$

k2 \ k1	1	2	3	4	5
1	1	0.98590	0.95258	0.91094	0.86730
2		1	0.98988	0.96616	0.93581
3			1	0.99283	0.97556
4				1	0.99472
5					1

6. МОМЕНТЫ СТАТИСТИКИ АНДЕРСОНА - ДАРЛИНГА

Статистика Андерсона- Дарлинга имеет вид [6] :

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{[S_n(x) - P(x)]^2}{P(x)\,[1 - P(x)]} dP(x) = n \int_0^1 \frac{[F_n(t) - t]^2}{t(1-t)} dt. \qquad (6.1)$$

Эта статистика является модификацией статистики омега- квадрат (ω_n^2), чувствительной к отклонениям на "хвостах" исследуемого распределения. Моменты распределения A_n^2 могут быть получены с использованием выражения для обобщенной моментной функции эмпирического процесса (1.10), где $\varphi[t, y_n(t)] = y_n(t) / [t(1 - t)]$.

Для определения дисперсии статистики A_n^2 была составлена программа на языке *muMATH*.

В отличие от рассмотренных выше случаев статистик $\omega(n,k)$ и $\omega_A(n,k)$ (см. гл. 4 и 5), для которых при определении моментов требовалось вычисление многократных интегралов от произведений полиномиальных функций, для моментов статистики A_n^2 необходимо вычисление таких интегралов от дробно- рациональных функций. Исходя из этого, при написании программы учитывались ограниченные возможности системы *muMATH*. В частности, было предусмотрено оперирование как с константами выражениями, соответствующими операциям интегрирования, которые система оказывалась не в состоянии выполнить.

В результате работы программы получены следующие результаты :

$$D(A_{n=1}^2) = 4 + 2\int_0^1 1/x \ln(1 - x)dx, \qquad E(A_{n=1}^2) = 1, \quad \text{для } n = 1,$$

$$D(A_{n=2}^2) = -1 - \int_0^1 1/x \ln(1 - x)dx, \qquad E(A_{n=2}^2) = 1, \quad \text{для } n = 2,$$

$$D(A^2_{n=3}) = -8/3 - 2\int_0^1 1/x \ln (1 - x)dx, \quad E(A^2_{n=3}) = 1, \quad \text{для } n = 3. \tag{6.2}$$

Интеграл, который не был вычислен самой системой *muMATH*, является табличным, и его значение равно

$$\int_0^1 1/x \ln (1 - x)dx = - \pi /6. \tag{6.3}$$

Используя выражения (6.2) и (6.3) для построения системы линейных уравнений относительно коэффициентов полинома, представляющего искомую аналитическую зависимость $D(A^2_n)$ от n, и решая ее, получаем окончательно

$$D(A^2_n) = \frac{30 - 3\pi^2 + 2n (\pi^2 - 9)}{3n}. \tag{6.4}$$

Из (6.4) следует, что в асимптотическом пределе ($n \longrightarrow \infty$)

$$D(A^2) = \frac{2 (\pi^2 - 9)}{3}. \tag{6.5}$$

Выражение (6.5) совпадает с вычисленным по формуле для дисперсии асимптотического распределения статистики W^2_n [6]:

$$\sigma^2 = 4 \int_0^1 (1 - s)^2 \psi (s) \int_0^s t^2 \psi(t)dtds,$$

где $W^2_n \equiv A^2_n$ при $\psi(x) = 1/[x(1 - x)]$.

7. ВЫЧИСЛЕНИЕ КОВАРИАЦИЙ И КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ СТАТИСТИК $\omega(n,k)$ И СТАТИСТИК A^2_n И ВАТСОНА (U^2_n)

Эта глава посвящена вычислениям некоторых "смешанных" ковариаций (C) и коэффициентов корреляции (ρ) основных интегральных статистик.

Ковариация и коэффициент корреляции статистик A^2_n и $\omega(n,k)$ с учетом $E [A^2_n] = 1$ записываются в виде

$$C [A^2_n \omega(n,k)] = E [A^2_n \omega(n,k)] - E [\omega(n,k)],$$

$$\rho = C [A^2_n \omega(n,k)] / \sqrt{D [A^2_n] D [\omega(n,k)]}. \tag{7.1}$$

Величины $E [\omega(n,k)]$, $D [\omega(n,k)]$ и $D [A^2_n]$ вычислялись ранее (см. гл. 4 и 6), и ,таким образом, выражения (7.1) содержат только один неизвестный параметр $E [A^2_n \omega(n,k)]$. Эта величина определяется выражением типа (3.5), в котором в качестве J_2 рассматривается обобщенная моментная функция (1.10) с $\varphi_1[t,y_n(t)] = y^2_n(t)/[t(1 - t)]$ и $\varphi_2[t',y_n(t')] = y^k_n(t')$.

Вычисление величин (7.1), как и ранее в гл.4,5 и 6, проводилось с использованием системы *muMATH* и последующим определением параметров искомого полинома как решения соответствующей системы линейных уравнений.

Приведем результаты для двух частных случаев $k = 2$ и $k = 4$.

а) $k = 2$

$$C [A^2_n \omega^2_n] = (2n - 1) / 18n. \tag{7.2}$$

$$\rho[A^2_n \omega^2_n] = (2n -1)/18n \sqrt{540n^2/((4n -3)(30 - 3\pi^2 + 2n(\pi^2 - 9))}. \tag{7.3}$$

В асимптотическом пределе при $n \longrightarrow \infty$ коэффициент корреляции (7.3) принимает вид

$$\rho [A^2\omega^2] = 1/9 \sqrt{135/(2(\pi^2 - 9))} \cong 0.9789228.$$

б) $k = 4$

$$C [A^2_n \omega^4_n] = (28 - 117n + 126n^2)/ 900n^2. \tag{7.4}$$

$$\rho [A^2_n \omega^4_n] = \sqrt{14} (28 - 117n + 126n^2)/$$

$$/(30 \sqrt{(188n^3 - 382n^2 + 276n - 70)(30 - 3\pi^2 + 2n(\pi^2 - 9))}. \tag{7.5}$$

В асимптотическом пределе при $n \longrightarrow \infty$ коэффициент корреляции (7.5) принимает вид

$$\rho [A^2\omega^4] = 21 \sqrt{14} / (10 \sqrt{94 (\pi^2 - 9)}) \cong 0.8690774 .$$

Наряду со статистикой Андерсона - Дарлинга широко используется так называемая статистика Ватсона, первоначально построенная для проверки гипотез о круговых распределениях и оказавшаяся более мощной, чем статистика ω^2_n, также и для некоторых задач на числовой прямой (см., например, [2,7]) :

$$U^2_n = n \int_0^1 [F_n(t) - t - \overline{F_n(t) - t}]^2dt = \int_0^1 y^2_n(t)dt - \overline{y}^2_n =$$

$$= \omega^2_n - n (\overline{t} - 1/2)^2, \qquad \text{где } \overline{t} = \sum_{i=1}^n t_i. \tag{7.6}$$

Для ковариации и коэффициента корреляции статистики Ватсона и омега-квадрат приведем только результаты вычислений

$$C [U^2_n\omega^2_n] = (7n - 2)/360n - 1/72 = (n - 1)/180n, \tag{7.7}$$

$$\rho [U^2_n\omega^2_n] = \sqrt{(2n - 2)/(4n - 3)} . \tag{7.8}$$

Из (7.8) видно, что при $n = 1$ коэффициент корреляции $\rho [U^2_n\omega^2_n] = 0$, а в асимптотическом пределе при $n \longrightarrow \infty$

$$\rho\ [U^2\omega^2] = \sqrt{1/2}\ .$$

Результат (7.8) совпадает с вычисленным ранее другим способом в работе [21].

8. АЛГЕБРАИЧЕСКАЯ ЗАВИСИМОСТЬ ОТ n НЕКОТОРЫХ МОМЕНТНЫХ ФУНКЦИЙ

На основании вида зависимостей (1.6), (1.7) и (1.8) сделаем предположения об алгебраическом представлении моментных функций J_q эмпирического процесса. Заметим, что индекс q определяет количество независимых аргументов $t^{I}, t^{II}, \dots, t^{(q)}$ моментной функции, поэтому для удобства назовем величину q размерностью функций типа (1.6). Общую методику определения алгебраического вида моментных функциий рассмотрим на примере функции $J_2(t_1,t_2;n,k_1,k_2)$ размерности $q = 2$.

Предположим : 1) $J_2(t_1,t_2;n,k_1,k_2)$ представляет собой произведение двух полиномов со степенями k_1 и k_2 по t без нулевых членов, 2) коэффициенты при членах этого произведения, в свою очередь, представляются полиномами заранее не известной степени по $1/\sqrt{n}$. Суммируем эти предположения в виде системы

$$\left\{\begin{aligned}
&J_2(t_1,t_2;\ n,k_1,k_2) = A_1t_1 + A_2t_1^2 + \dots + A_{k_1}t_1^{k_1} + B_1t_2 + B_2t_2^2 + \\
&\quad + \dots + B_{k_2}t_2^{k_2} + C_1t_1t_2 + C_2t_1^2t_2 + C_3t_1t_2^2 + \dots + C_{k_1k_2}t_1^{k_1}\ t_2^{k_2}\ ,\\
&A_1 = a_1 + b_1/\sqrt{n} + c_1/n + \dots ,\\
&A_2 = a_2 + b_2/\sqrt{n} + c_2/n + \dots ,\\
&\dots\\
&A_{k_1} = a_{k_1} + b_{k_1}/\sqrt{n} + c_{k_1}/n + \dots ,\\
&B_1 = a_1' + b_1'/\sqrt{n} + c_1'/n + \dots ,\\
&B_2 = a_2' + b_2'/\sqrt{n} + c_2'/n + \dots ,\\
&\dots\\
&B_{k_2} = a_{k_2}' + b_{k_2}'/\sqrt{n} + c_{k_2}/n + \dots ,\\
&C_1 = a_1'' + b_1''/\sqrt{n} + c_1''/n + \dots ,\\
&C_2 = a_2'' + b_2''/\sqrt{n} + c_2''/n + \dots ,\\
&\dots\\
&C_{k_1k_2} = a_{k_1k_2}'' + b_{k_1k_2}''/\sqrt{n} + c_{k_1k_2}''/n + \dots .
\end{aligned}\right. \tag{8.1}$$

Для каждого из коэффициентов A_i, $i = 1, 2, \dots, k_1$; B_j, $j = 1, 2, \dots, k_2$, k_2; C_s, $s = 1, 2, \dots, k_1k_2$ составляется система с заранее не известным числом уравнений (подбирается эмпирически) и решается относительно a_1, b_i, ... ; a_i', b_i', ... ; a_1'', b_1'', ... Восстановленный таким образом вид $J_2(t_1,t_2;n,k_1,k_2)$ далее проверяется на соответствие значениям функции (1.6) при $q = 2$ для новых n.

Очевидно, что систему типа (8.1) можно составить для J_q любой размерности.

Остановимся на некоторых результатах

$$1)\ q = 1,\quad J_1(t;\ n,k) = E\left\{\ y_n^k(t)\right\}.$$

$$J_1(t;\ n,1) = 0, \qquad J_1(t;\ n,2) = t - t^2. \tag{8.2}$$

$$J_1(t;\ n,3) = 1/\sqrt{n}\ (t - 3\ t^2 + 2\ t^3) = 1/\sqrt{n}\ t(1 - t)(1 - 2t). \tag{8.3}$$

$$J_1(t;n,4) = 1/n\ t(1 - t)(1 - 2t)(1 - 3t) - 1/n\ t^2(1 - t) + 3\ t^2(1 - t)^2. \tag{8.4}$$

$$J_1(t;\ n,5) = 10/\sqrt{n}\ t^2(1 - 4\ t + 5\ t^2 - 2\ t^3) + 1/n^{3/2}\ t\ (1 - 15\ t + 50\ t^2 - 60\ t^3 + 24\ t^4). \tag{8.5}$$

$$J_1(t;\ n,6) = 15t_1^3(\ 1 - t_1)^3 + 1/n\ [\ 5t_1^2\ (\ 5 - 36t_1 + 83t_1^2 - 78t_1^3 + 26t_1^4)] + 1/n^2\ [\ t_1(\ 1 - 31\ t_1 + 180\ t_1^2 - 390\ t_1^3 + 360\ t_1^4 - 120\ t_1^5\ )]. \tag{8.6}$$

Поскольку среднее значение статистики $\omega(n,k)$ связано с функцией $J_1(t;\ n,k)$ равенством (см. гл.2)

$$E\ [\omega(n,k)] = \int_0^1 J_1(t;\ n,k)dt, \tag{8.7}$$

то подтверждением истинности результатов (8.2) - (8.6) может являться установление их связи с выражениями для средних значений (4.3) и (4.5) посредством (8.7).

$$2)\ q = 2,\quad J_2(t_1,t_2;\ n,k_1,k_2) = E\left\{\ y_n^{k_1}(t_1)y_n^{k_2}(t_2)\right\}.$$

Рассмотрим результаты для k_1, $k_2 = 1,2,3$:

a) $J_2(t_1,t_2;\ n,1,1) = t_1 - t_1t_2$, при $t_1 < t_2$,

$J_2(t_1,t_2;\ n,1,1) = t_2 - t_1t_2$, при $t_1 > t_2$,

или в общем виде

$$J_2(t_1,t_2;\ n,1,1) = E\left\{ y_n(t_1) y_n(t_2) \right\} = \min(t_1,t_2) - t_1 t_2, \qquad (8.8)$$

что является хорошо известным результатом для функции ковариации эмпирического процесса (см., например, [2]).

б) $J_2(t_1,t_2;\ n,1,2) = 1/\sqrt{n}\ t_1(1 - t_2)(1 - 2t_2)$, при $t_1 < t_2$,

$J_2(t_1,t_2;\ n,1,2) = 1/\sqrt{n}\ t_2(1 - t_1)(1 - 2t_2)$, при $t_1 > t_2$.

В асимптотическом пределе $n \longrightarrow \infty$ значение этой функции стремится к 0.

в) $$J_2(t_1,t_2;\ n,2,2) = t_1^2(1 - t_2)(2 - 3t_2) + t_1 t_2(1 - t_2) + 1/n\ t_1(1 - t_2)(1 - 2t_2) - 2/n\ t_1^2(1 - t_2)(2 - 3t_2)$$
при $t_1 < t_2$. $\qquad (8.9)$

Для случая $t_1 < t_2$ необходимо в этой формуле t_1 и t_2 поменять местами.

г) $$J_2(t_1,t_2;\ n,1,3) = 1/n\ t_1(1 - t_2)(1 - 2t_2)(1 - 3t_2) - 1/n\ t_1 t_2(1 - t_2) + 3\ t_1 t_2(1 - t_2)^2,$$ при $t_1 < t_2$.

$$J_2(t_1,t_2;\ n,1,3) = 1/n\ t_2(1 - t_1) - 6/n\ t_2^2(1 - t_2) \times (1 - t_1) + 3\ t_2^2(1 - t_2)(1 - t_1),$$ при $t_1 > t_2$.

д) $$J_2(t_1,t_2;\ n,2,3) = 1/\sqrt{n}\,(6t_1^2 + 4\ t_1 t_2 - 31\ t_1^2 t_2 - 9\ t_1 t_2^2 + 45\ t_1^2 t_2^2 + 5\ t_1 t_2^3 - 20\ t_1^2 t_2^3) + 1/n^{3/2}\ (t_1 - 8\ t_1^2 - 7\ t_1 t_2 + 38\ t_1^2 t_2 + 12\ t_1 t_2^2 - 54\ t_1^2 t_2^2 - 6\ t_1 t_2^3 + 24\ t_1^2 t_2^3),$$ при $t_1 < t_2$.

$$J_2(t_1,t_2;\ n,2,3) = 1/\sqrt{n}\ (9\ t_2^2 - 15\ t_2^3 + t_2 t_1 - 24\ t_2^2 t_1 - t_2 t_1^2 + 15\ t_2^2 t_1^2 + 35\ t_2^3 t_1 - 20\ t_2^3 t_1^2) + 1/n^{3/2}\ (t_2 - 12\ t_2^2 + 18\ t_2^3 - 3\ t_2 t_1 + 30\ t_2^2 t_1 + 2\ t_2 t_1^2 - 18\ t_2^2 t_1^2 - 42\ t_2^3 t_1 + 24\ t_2^3 t_1^2),$$ при $t_1 > t_2$. $\qquad (8.10)$

е) $$J_2(t_1,t_2;n,3,3) = 6\ t_1^3 + 9\ t_1^2 t_2 - 18\ t_1^2 t_2^2 - 27\ t_1^3 t_2 + 36\ t_1^3 t_2^2 + 9\ t_1^2 t_2^3 - 15\ t_1^3 t_2^3 + 1/n\ (21\ t_1^2 - 57\ t_1^3 + 4\ t_1 t_2 - 114\ t_1^2 t_2 - 9\ t_1 t_2^2 + 171\ t_1^2 t_2^2 + 239\ t_1^3 t_2 + 5\ t_1 t_2^3 - 312\ t_1^3 t_2^2 - 78\ t_1^2 t_2^3 +$$

$$+ 130\ t_1^3 t_2^3) + 1/n^2\ (t_1 - 24\ t_1^2 + 54\ t_1^3 - 7\ t_1 t_2 + 114\ t_1^2 t_2 + 12\ t_1 t_2^2 - 162\ t_1^2 t_2^2 - 222\ t_1^3 t_2 - 6\ t_1 t_2^3 + 288\ t_1^3 t_2^2 + 72\ t_1^2 t_2^3 - 120\ t_1^3 t_2^3)$$ при $t_1 < t_2$. $\qquad (8.11)$

Для случая $t_1 > t_2$ необходимо в формуле (8.11) t_1 и t_2 поменять местами.

Моментные функции $J_2(t_1,t_2;\ n,k_1,k_2)$ на основании изложенного в предыдущих главах легко могут быть связаны с дисперсиями и ковариациями статистик $\omega(n,k)$, полученными ранее (см. формулы (3.7),(4.1) и (4.2)). Это дает возможность проверки выражений (8.8) - (8.11).

3) $q = 3$, $J_3(t_1,t_2,t_3;\ n,k_1,k_2,k_3) = E\left\{ y_n^{k_1}(t_1) y_n^{k_2}(t_2) y_n^{k_3}(t_3) \right\}$.

$$J_3(t_1,t_2,t_3;n,1,1,1) = 1/\sqrt{n}\,(t_1 - 2t_1 t_2 - t_1 t_3 + 2t_1 t_2 t_3)$$ при $t_1 < t_2 < t_3$. $\qquad (8.12)$

$$J_3(t_1,t_2,t_3;n,2,2,2) = \alpha_1 t_1 + \alpha_2 t_1^2 + \alpha_3 t_1 t_2 + \alpha_4 t_1 t_3 + \alpha_5 t_1^2 t_2 + \alpha_6 t_1^2 t_3 + \alpha_7 t_1 t_2^2 + \alpha_8 t_1 t_3^2 + \alpha_9 t_1^2 t_2^2 + \alpha_{10} t_1^2 t_3^2 + \alpha_{11} t_1 t_2 t_3 + \alpha_{12} t_1^2 t_2 t_3 + \alpha_{13} t_1 t_2^2 t_3 + \alpha_{14} t_1 t_2 t_3^2 + \alpha_{15} t_1^2 t_2^2 t_3 + \alpha_{16} t_1^2 t_2 t_3^2 + \alpha_{17} t_1 t_2^2 t_3^2 + \alpha_{18} t_1^2 t_2^2 t_3^2$$ при $t_1 < t_2 < t_3$, где

$\alpha_1 = 1/n$, $\alpha_2 = 14/n - 16/n^2$, $\alpha_3 = 10/n - 12/n^2$, $\alpha_4 = 1/n - 3/n^2$,
$\alpha_5 = 10 - 95/n + 90/n^2$, $\alpha_6 = 2 - 38/n + 40/n^2$, $\alpha_7 = 2 - 19/n + 18/n^2$,
$\alpha_8 = - 1/n + 2/n^2$, $\alpha_9 = -12 + 104/n - 96/n^2$, $\alpha_{10} = -2 + 24/n - 24/n^2$,
$\alpha_{11} = 1 - 27/n + 30/n^2$, $\alpha_{12} = -25 + 225/n - 210/n^2$, $\alpha_{13} = -5 + 45/n - 42/n^2$, $\alpha_{14} = -1 + 17/n - 18/n^2$, $\alpha_{15} = 27 - 234/n + 216/n^2$,
$\alpha_{16} = 15 - 130/n + 120/n^2$, $\alpha_{17} = 3 - 26/n + 24/n^2$,
$\alpha_{18} = -15 + 130/n - 120/n^2$. $\qquad (8.13)$

3) $q = 4$.

С увеличением размерности q резко возрастают вычислительные трудности при определении алгебраического вида соответствующих моментных функций, а получаемые выражения становятся весьма громоздкими. Так, для вычисления $J_4(t_1,t_2,t_3,t_4;\ n,2,2,2,2)$ потребовалось определить рассмотренным выше способом 54 коэффициента, имеющих структуру $(a + bn + cn^2 + dn^3)/n^3$.

Явный вид J_4 наряду с выражением для J_3 (8.13) использовались для вычисления третьего и четвертого моментов статистики омега-квадрат по формулам

$$M_3[\omega(n,2)] = 3!\int_0^1 dt_3 \int_0^{t_3} dt_2 \int_0^{t_2} dt_1\ J_3(t_1< t_2< t_3;\ n,2,2,2).$$

$$M_4[\omega(n,2)] = 4!\int_0^1 dt_4 \int_0^{t_4} dt_3 \int_0^{t_3} dt_2 \int_0^{t_2} dt_1\ J_4(t_1<t_2<t_3<t_4\ ;\ n,2,2,2,2).$$

Выражения для $M_3[\omega(n,2)]$ и $M_4[\omega(n,2)]$, в свою очередь, использовались для определения центральных моментов соответствующих порядков :

$$\mu_3 = \frac{32\ n^2 - 61\ n + 30}{3780\ n^2}\ , \qquad (8.14)$$

$$\mu_4 = \frac{496\ n^3 - 1532\ n^2 + 1671\ n - 630}{75600\ n^3}\ . \qquad (8.15)$$

Формулы (8.14) и (8.15) полностью совпадают с вычисленными непосредственно из определения соответствующих моментов в работе [22].

9. МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЙ $\omega(n,k)$, $\omega_A(n,k)$, A_n^2 И U_n^2

Для дополнительной проверки результатов, полученных в предыдущих главах, по определению моментов, корреляционных функций и коэффициентов корреляции для указанных непараметрических статистик было проведено моделирование их распределений. С этой целью для некоторых значений объема выборки n были написаны программы на фортране для следующих пар статистик :

а) $\omega(n,2)$ и $\omega(n,4)$, б) $\omega(n,2)$ и U_n^2, в) $\omega(n,2)$ и A_n^2 ,

г) $\omega(n,4)$ и A_n^2, д) $\omega_A(n,1)$ и $\omega_A(n,2)$,

причем число разыгранных значений для каждой программы составляло 20000. При моделировании распределения $\omega(n,k)$ использовалась формула алгебраического представления статистики $\omega(n,k)$*:

---

*) Вопросы, связанные с алгебраическим видом статистик $\omega(n,k)$ и пределами их изменений, будут рассмотрены позднее, в работе , посвященной изучению функций распределения этих статистик и построению критериев согласия (совместно с В.В.Ивановым).

$$\omega(n,k) = - n^{k/2}/(k+1)\sum_{i=1}^{n}\left\{\left((i-1)/n - t_i\right)^{k+1} - \left(i/n - t_i\right)^{k+1}\right\}. \qquad (9.1)$$

Вывод формулы алгебраического вида статистики $\omega_A(n,k)$ дан в Приложении :

$$\omega_A(n,\ k) = n^{k/2}\ /(\ k+1\ )\sum_{i=1}^{n}\left\{\left(i/n - t_i\right)^{k+1} sign^k(i/n - t_i) - \right.$$

$$\left. - \left((i-1)/n - t_i\right)^{k+1} sign^k\left((i-1)/n - t_i\right)\right\}. \qquad (9.2)$$

Ниже приведены результаты моделирования для первого и второго моментов соответствующих статистик, а также их ковариации (дисперсии и коэффициент корреляции вычисляются на основе этих величин). Погрешности результатов моделирования представлены в виде оценок среднеквадратичных отклонений. Объем выборки $n = 2$.

а) $E\ [\omega(n,2)] = 0.16616 \pm 0.00083$, $E([\omega(n,2)]^2) = 0.04148 \pm 0.00044$,
$E\ [\omega(n,4)] = 0.08297 \pm 0.00082$, $E([\omega(n,4)]^2) = 0.02048 \pm 0.00041$,
$D\ [\omega(n,2)] = 0.01388$, $D\ [\omega(n,4)] = 0.01360$,
$cov\ [\omega(n,2)\omega(n,4)] = 0.01327$, $\rho = 0.96635$;

б) $E\ [U_n^2] = 0.08309 \pm 0.00026$, $E([U_n^2]^2) = 0.00828 \pm 0.00005$,
$D\ [U_n^2] = 0.00138$,
$cov\ [\omega(n,2)\ U_n^2] = 0.00278$, $\rho = 0.63460$;

в) $E\ [A_n^2] = 0.99954 \pm 0.00571$, $E([A_n^2]^2) = 1.65025 \pm 0.07595$,
$D\ [A_n^2] = 0.65118$,
$cov\ [\omega(n,2)\ A_n^2] = 0.08330$, $\rho = 0.87628$;

г) $cov\ [\omega(n,4)\ A_n^2] = 0.08264$, $\rho = 0.87817$;

д) $E\ [\omega_A(n,1)] = 0.32352 \pm 0.00081$, $E([\omega_A(n,1)]^2) = 0.11769 \pm 0.00061$,
$E\ [\omega_A(n,2)] = 0.16616 \pm 0.00083$, $E([\omega_A(n,2)]^2) = 0.04149 \pm 0.00044$,
$D\ [\omega_A(n,1)] = 0.01302$, $D\ [\omega_A(n,2)] = 0.01388$,
$cov\ [\omega_A(n,1)\omega_A(n,2)] = 0.01325$, $\rho = 0.98585$;

Сравнение этих величин со значениями табл.3 или со значениями, вычисленными по формулам, приведенным в главах 3 - 7, показывает, что результаты моделирования совпадают в пределах статистической ошибки с результатами вычислений по аналитическим формулам.

ЗАКЛЮЧЕНИЕ

В рамках формального определения обобщенной интегральной непараметрической статистики, представимой в виде следующего функционала от эмпирического процесса $y_n(t)$:

$$\omega_n(\varphi) = \int_0^1 \varphi\,[t, y_n(t)]dt,$$

где φ - некоторая функция этого процесса, рассмотрены статистики

$$\omega(n,k) = \int_0^1 y_n^k(t)dt \qquad \text{и} \qquad \omega_A(n,k) = \int_0^1 |y_n(t)|^k dt,$$

а также хорошо известные статистики - Мозеса $\bar{U}_n$, Ватсона (U_n^2), Андерсона- Дарлинга (A_n^2) и статистики ω_n^2 и ω_n^3 , подпадающие под определение $\omega(n,k)$ соответственно при $k = 2$ и $k = 3$.

Рассмотрен общий метод исследования моментных и корреляционных свойств статистик типа $\omega_n(\varphi)$, основанный на свойствах моментных функций процесса $y_n(t)$.

Определен вид обобщенной моментной функции эмпирического процесса J_q, а также аналитические зависимости от n для ряда ее конкретных реализаций :

а) $q = 1$, $\varphi\,[t, y_n(t)] = [y_n(t)]^k$, $k = 1,2,\ldots,6$.

б) $q = 2$, $\varphi_i[t, y_n(t)] = [y_n(t)]^{k_i}$, $k_i = 1,2,3$, i ; 1,2.

в) $q = 3$, $\varphi_j[t, y_n(t)] = [y_n(t)]^2$, $j = 1,2,3$.

г) $q = 4$, $\varphi_i[t, y_n(t)] = [y_n(t)]^2$, $i = 1, 2, 3, 4$.

Получены выражения для ковариаций и коэффициентов корреляции статистик $\omega(n,k_1)$ и $\omega(n,k_2)$ для k_1, $k_2 = 1,2,3,4$ при произвольном значении n. Установлена сильная положительная коррелированность этих статистик при четном $|\, k_1 - k_2 |$.

Для ранее известных статистик $\bar{U}_n$, U_n^2, ω_n^2, A_n^2 и ω_n^3 получены новые результаты. В частности - выражения, определяющие зависимость от объема

эмпирической выборки n дисперсии распределения A_n^2, а также коэффициентов корреляции этих статистик.

ПРИЛОЖЕНИЕ. АЛГЕБРАИЧЕСКИЙ ВИД СТАТИСТИКИ $\omega_A(n,k)$.

Рассмотрим интегральный вид статистики $\omega_A(n,k)$ (см. гл. 6)

$$\omega_A(n,\ k) = n^{k/2}\int_0^1 |\ F_n(t) - t\ |^k dt = n^{k/2}\int_0^{t_1} t^k dt +$$

$$+ n^{k/2}\sum_{i=1}^{n-1}\int_{t_i}^{t_{i+1}} |\ i/n - t\ |^k dt + n^{k/2}\int_{t_n}^{1} |\ 1 - t\ |^k dt\ . \tag{1}$$

Запишем через I_i, $i = 1,2,3$ соответствующие члены (1), разделенные знаками " + ".

$$I_1 = n^{k/2}\ t_1^{k+1} /(\ k + 1\ ), \qquad I_3 = n^{k/2}\ (\ 1 - t_n)^{k+1} /(\ k + 1\ ),$$

$$I_2 = -\ n^{k/2}\sum_{i=1}^{n-1}\int_{i/n-t_i}^{i/n-t_{i+1}} |\ x\ |^k dx = -\ n^{k/2}\sum_{i=1}^{n-1}\int_{a_i}^{b_i} |\ x\ |^k dx\ , \tag{2}$$

где $a_i = i/n - t_i$, $b_i = i/n - t_{i+1}$. Рассмотрев различные варианты соотношений между a_i и b_i и нулем, получим

$$\int_{a_i}^{b_i} |\ x\ |^k dx = -\ 1/(\ k + 1\ )\Big\{\ a_i^{k+1}\big(sign(\ a_i)\big)^k - b_i^{k+1}\ \big(\ sign(\ b_i)\big)^k\Big\}\ . \tag{3}$$

На основании (2) и (3), таким образом, получаем

$$I_2 = n^{k/2}\sum_{i=1}^{n-1} 1/(\ k + 1\ )\ \Big[\ \big(i/n - t_i\big)^{k+1}\ sign^k(\ i/n - t_i) -$$

$$-\ (\ i/n - t_{i+1})^{k+1}\ sign^k(\ i/n - t_{i+1})\ \Big] = n^{k/2} /(\ k + 1\ )\ \times$$

$$\times\ \Big(\ \sum_{i=1}^{n}(\ i/n - t_i\ )^{k+1}\ sign^k(\ i/n - t_i) - \sum_{j=1}^{n}[\ (\ j - 1\ )/n - t_j\ ]^{k+1}\ \times$$

$$\times\ sign^k[(\ j - 1\ )/n - t_j\ ] - (\ 1 - t_n)^{k+1} + \underbrace{(-1)^{2k+1}}_{=-1} t_1^{k+1}\Big)\ .$$

Таким образом, суммируя I_1, I_2 и I_3 , получаем

$$\omega_A(n,\ k) = n^{k/2} /(\ k + 1\ )\sum_{i=1}^{n}\Big\{\ (\ i/n - t_i)^{k+1} sign^k(\ i/n - t_i) -$$

$$- [(i-1)/n - t_i]^{k+1} \mathrm{sign}^k [(i-1)/n - t_i] \Big\}. \qquad (4)$$

Путем обобщения частных результатов получено выражение для максимального значения статистики $\omega_A(n,k)$

$$[\omega_A(n,k)]_{max} = n^{k/2} / (k+1). \qquad (5)$$

Выборочные значения $t_i = (2i-1)/2n$, $i = 1,2,\dots,n$ минимизируют значение статистики $\omega_A(n,k)$, поэтому ее минимальное значение равно

$$[\omega_A(n,k)]_{min} = n^{k/2}/(k+1) \sum_{i=1}^{n} \left\{ \left((i-1)/n - (2i-1)/2n \right)^{k+1} - \right.$$

$$\left. - \left(i/n - (2i-1)/2n \right)^{k+1} \right\} = 1/(2^k n^{k/2} (k+1)). \qquad (6)$$

ЛИТЕРАТУРА

1. *Gaenssler P. and Stute W. "Empirical Processes : A Survey of Results for independent and Identically Distributed Random Variables". The Annals of Probability. 1979, vol.7, no 2, p.193-243.*
2. Мартынов Г.В. Критерии омега- квадрат. М.: Наука, *1978.*
3. Смирнов Н.В., Дунин- Барковский И.В. Курс теории вероятностей и математической статистики. М.: Наука, *1965.*
4. Зрелов П.В., Иванов В.В. "Проверочная статистика $\omega_n^3 = n^{3/2}\int_{-\infty}^{\infty} [S_n(x) - P(x)]^3 dP(x)$ в теории непараметрических критериев согласия ", сообщение ОИЯИ *Р10-88-321*, Дубна, *1988.*
5. Зрелов П.В., Иванов В.В. Сообщение ОИЯИ Р*10-89-577*, Дубна, *1989.*
6. *Anderson T.W., Darling D.A.- Ann. Math. Statist., 1952, v.23, p.193.*
7. *Durbin J. Distribution Theory for Tests Based on the Sample Distribution Function . Society for Industrial and Applied Mathematics, Philadelphia, Pensylvania 19103, 1973.*
8. Соболь И.М. Численные методы Монте- Карло. М.: Наука, *1973.*
9. *Mises R. von. Warscheinlichkeitsrechnung und ihre Anvedungen in der Statistik und Phisik. Leipzig; Wien: Franz Deuticke, 1931,p.574.*
10. *Cramer H. On the composition of elementary errors. Second paper: statistical applications.- Skand.Akturietidskr., 1928, h.1/2, p.141-180.*
11. *G. Inst. ital. attuari, 1933, vol.4, N1.*
Kolmogorov A.N. Sulla Determinazione Empirica di una legge di distribuzione. p.83-91.
Glivenko V. Sulla determinazione empirica delle leggi di probabilita. p.92-99.
12. Смирнов Н.В. О распределении ω^2- критерия Мизеса.- Мат.сб., *1937*, т.*2*, N .*1*, с.*973-993.*
13. Боровков А.А. Теория вероятностей. Москва : Наука , *1982.*
14. Смирнов Н.В. О критерии Крамера- Мизеса. УМН, *1949*, т.*4*, вып.*4*, с.*196-197.*
15. *Marshall A.W. The small Sample Distribution* $n\omega_n^2$ *, Ann. Math. Statist., 1958, v.29, p.307.*
16. *Stephens M., Maag U.R. Futher Percentage Points for* W^2*, Biometrica, 1968,v.55, p.428.*
17. Яноши Л. Теория и практика обработки результатов измерений. Москва: Мир , *1965.*
18. Гихман И.И., Скороход А.В. Введение в теорию случайных процессов. Москва: Наука , *1977.*
19. Королюк В.С., Портенко Н.И., Скороход А.В., Турбин А.Ф. Справочник по теории вероятностей и математической статистике. Москва: Наука , *1985.*
20. *Stephens M.A. The distribution of the goodness-of-fit statistic,* U_n^2 *.I, Biometrica, 1963, v.50, p.303.*
21. *Pearson E.S. Comparison of tests for randomness of points on a line, Biometrika, 1963, v.50, 3 and 4, p.315.*
22. *Pearson E.S., Stephens M.A. The goodness-of-fit tests based on* W_n^2 *and* U_n^2 *, Biometrika, 1962, 49, 3 and 4, p.397.*
23. Королюк В.С., Боровских Ю.В., Аналитические проблемы асимптотики вероятностных распределений. Киев: Наукова думка , *1981.*
24. *Birnbaum Z.W., Tang V.K.T.- Rev. Inst. Statist., 1964, 32, p.2.*

Рукопись поступила в издательский отдел
29 сентября 1992 года.

Зрелов П.В. Р11-92-398

Обобщенные моментные функции эмпирического процесса и интегральные непараметрические статистики

Предложена и исследована обобщенная интегральная непараметрическая статистика, представимая в виде следующего функционала от эмпирического процесса $y_n(t)$: $\omega_n(\varphi) = \int_0^1 \varphi\,[y_n(t)]\,dt$. Рассмотрен общий метод исследования статистик такого типа, основанный на свойствах моментных функций процесса $y_n(t)$. Определен вид обобщенной моментной функции эмпирического процесса J_φ, а также аналитические зависимости от n для ряда ее конкретных реализаций. В рамках $\omega_n(\varphi)$ рассмотрены статистики $\omega\,(n, k) = \int_0^1 y_n^k\,(t)dt$ и $\omega_A(n, k) = \int_0^1 |\,y_n\,(t)\,|^k dt$, а также хорошо известные статистики Мозеса, Ватсона, Андерсона — Дарлинга и омега-квадрат. Получены выражения, определяющие моменты распределений, ковариации и коэффициенты корреляции статистик $\omega(n, k_1)$ и $\omega(n, k_2)$, а также статистик U_n, U_n^2, ω_n^2, A_n^2 и ω_n^3 для ряда значений k_1 и k_2.

Работа выполнена в Лаборатории вычислительной техники и автоматизации ОИЯИ.

Сообщение Объединенного института ядерных исследований. Дубна 1992

Перевод автора

Zrelov P.V. P11-92-398

The Generalized Moment Functions of Empirical Process and the Integral Nonparametric Statistics

The generalized nonparametric statistics represented as a following functional of the empirical process $y_n(t)$ $\omega_n(\varphi) = \int_0^1 \varphi\,[y_n(t)]\,dt$ was proposed and investigated. The general method of the investigation of the statistics of such a type based on the properties of the moment functions of $y_n(t)$ has been considered. The form of the generalized moment function of the empirical process J_q as well as the analytical dependences upon n for the series of its concrete realizations was determined. The statistics $\omega\,(n, k) = \int_0^1 y_n^k\,(t)dt$ and $\omega_A(n, k) = \int_0^1 |\,y_n\,(t)\,|^k dt$, and the well known Moses, Watson, Anderson — Darling and omega-squared statistics has been considered in frames of $\omega_n(\varphi)$. The expressions for the distribution moments, covariances and correlation coefficients of the $\omega(n, k_1)$ and $\omega(n, k_2)$ statistics as well as U_n, U_n^2, ω_n^2, A_n^2 and ω_n^3 statistics has been performed for the series of k_1 and k_2 values.

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

Communication of the Joint Institute for Nuclear Research. Dubna 1992