УДК 519.2:303.732.4

ПРЕДЕЛЬНАЯ ТЕОРИЯ НЕПАРАМЕТРИЧЕСКИХ СТАТИСТИК

Орлов Александр Иванович
д.э.н., д.т.н., к.ф.-м.н., профессор

Московский государственный технический университет им. Н.Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., 5, prof-orlov@mail.ru

Изучено асимптотическое поведение широкого класса непараметрических статистик, в который входят статистики типа омега-квадрат и типа Колмогорова-Смирнова. Доказаны предельные теоремы. Разработан метод аппроксимации ступенчатыми функциями, с его помощью получен ряд необходимых и достаточных условий

Ключевые слова: МАТЕМАТИЧЕСКАЯ СТАТИСТИКА, ПРЕДЕЛЬНЫЕ ТЕОРЕМЫ, СТАТИСТИКИ ИНТЕГРАЛЬНОГО ТИПА, НЕОБХОДИМЫЕ И ДОСТАТОЧНЫЕ УСЛОВИЯ, НЕПАРАМЕТРИЧЕСКИЕ СТАТИСТИКИ, СТАТИСТИКИ ТИПА ОМЕГА-КВАДРАТ, СТАТИСТИКИ ТИПА КОЛМОГОРОВА - СМИРНОВА

UDC 519.2:303.732.4

LIMIT THEORY OF NONPARAMETRIC STATISTICS

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci., professor
Bauman Moscow State Technical University, Moscow, Russia

We have studied the asymptotic behavior of a broad class of nonparametric statistics, which includes statistics of omega-square type and Kolmogorov-Smirnov type. Limit theorems have been proved. We have also developed the method of approximation with step functions. With the help of this method we have obtained a number of necessary and sufficient conditions

Keywords: MATHEMATICAL STATISTICS, LIMIT THEOREMS, INTEGRAL TYPE STATISTICS, NECESSARY AND SUFFICIENT CONDITIONS, NONPARAMETRIC STATISTICS, STATISTICS OF OMEGA-SQUARE TYPE, STATISTICS OF KOLMOGOROV — SMIRNOV TYPE

1. Введение

В прикладной статистике широко используются статистики типа омега-квадрат и типа Колмогорова - Смирнова [1]. Они применяются для проверки согласия с фиксированным распределением или семейством распределений, для проверки однородности двух выборок, симметрии распределения относительно 0, при оценивании условной плотности и регрессии в пространствах произвольной природы и т.д. В настоящей статье изучено асимптотическое поведение широкого класса непараметрических статистик, в который входят статистики типа омега-квадрат и типа Колмогорова-Смирнова. Доказаны предельные теоремы. Разработан метод аппроксимации ступенчатыми функциями, с его помощью получен ряд необходимых и достаточных условий. Доказательства теорем впервые публикуются в научной периодике.

2. Статистики интегрального типа и их асимптотика

Рассмотрим введенное автором настоящей статьи естественное обобщение статистик типа омега-квадрат - статистики интегрального типа

$$\xi_\alpha = \xi(f_\alpha, F_\alpha) = \int_X f_\alpha(x,\omega) dF_\alpha(x,\omega), \qquad (1)$$

где X – некоторое пространство, по которому происходит интегрирование (например, X = [0; 1], $X = R^1$ или $X = R^k$). Здесь {α} – направленное множество, переход к пределу по которому обозначен как α→∞ (в терминологии общей топологии [2]). Случайные функции f_α: $X\times\Omega \to Y$ обычно принимают значения, являющиеся числами. Но иногда рассматривают и постановки, в которых $У = R^k$ или $У$ – банахово пространство (т.е. полное нормированное пространство [3]). Наконец, $F_\alpha(x,\omega)$ – случайная функция распределения или случайная вероятностная мера; в последнем случае используют также обозначение $dF_\alpha(x,\omega)$= $F_\alpha(dx,\omega)$.

Предполагаются выполненными необходимые для корректности выводов внутриматематические предположения измеримости, например, сформулированные в [4, 5].

Пример 1. Рассмотрим критерий Лемана – Розенблатта, т.е. критерий типа омега-квадрат для проверки однородности двух независимых выборок (см. [6, гл. 5]). Его статистика имеет вид:

$$A = \frac{mn}{m+n}\int_{-\infty}^{\infty}(F_m(x) - G_n(x))^2 dH_{m+n}(x) ,$$

где $F_m(x)$ – эмпирическая функция распределения, построенная по первой выборке объема m, $G_n(x)$) - эмпирическая функция распределения, построенная по второй выборке объема n, а $H_{m+n}(x)$ – эмпирическая функция распределения, построенная по объединенной выборке объема $m+n$. Легко видеть, что

$$H_{m+n}(x) = \frac{m}{m+n} F_m(x) + \frac{n}{m+n} G_n(x).$$

Ясно, что статистика A имеет вид (1). При этом x – действительное число, X = $У$ = R^1, в роли α выступает пара (m, n), и $\alpha\to\infty$ означает, что $\min(m, n) \to \infty$. Далее,

$$f_\alpha(x,\omega) = \frac{mn}{m+n}\left(F_m(x) - G_n(x)\right)^2.$$

Наконец, $F_\alpha(x,\omega) = H_{m+n}(x)$.

Теперь обсудим асимптотическое поведение функций $f_\alpha(x,\omega)$ и $F_\alpha(x,\omega)$, с помощью которых определяется статистика A. Ограничимся случаем, когда справедлива гипотеза однородности, функции распределения, соответствующие генеральным совокупностям, из которых взяты выборки, совпадают. Их общую функцию распределения обозначим $F(x)$. Она предполагается непрерывной.

Введем в рассмотрение выборочные процессы

$$\xi_m(x) = \sqrt{m}(F_m(x) - F(x)),\ \ \eta_n(x) = \sqrt{n}(G_n(x) - F(x)).$$

Нетрудно проверить, что

$$f_\alpha(x,\omega) = \left(\sqrt{\frac{n}{m+n}}\xi_m(x) - \sqrt{\frac{m}{m+n}}\eta_n(x)\right)^2.$$

Сделаем замену переменной t = $F(x)$. Тогда выборочные процессы переходят в соответствующие эмпирические процессы (см. [7, разд. П5]):

$$f_\alpha(F^{-1}(t),\omega) = \left(\sqrt{\frac{n}{m+n}}\xi_m(t) - \sqrt{\frac{m}{m+n}}\eta_n(t)\right)^2,\ 0 \le t \le 1.$$

Конечномерные распределения этого процесса, т.е. распределения случайных векторов

$$(f_\alpha(F^{-1}(t_1),\omega), f_\alpha(F^{-1}(t_2),\omega),..., f_\alpha(F^{-1}(t_k),\omega))$$

для всех возможных наборов (t_1, t_2, … , t_k), сходятся к конечномерным распределениям квадрата броуновского моста $\xi^2(t)$. В соответствии с [7, разд. П5] рассматриваемая сходимость по распределению обозначается так:

$$f_\alpha(F^{-1}(t_1),\omega) \Rightarrow \xi^2(t),\ 0 \le t \le 1. \qquad (2)$$

Нетрудно видеть, что

$$F_\alpha(x,\omega) = H_{m+n}(x) \to F(x)$$

при $\alpha\to\infty$. С помощью замены переменной $t = F(x)$ получаем, что

$$F_\alpha(F^{-1}(t),\omega) = H_{m+n}(F^{-1}(t)) \to t \qquad (3)$$

при $\alpha\to\infty$. Из соотношений (2) и (3) хотелось бы сделать вывод, что в случае статистики Лемана - Розенблатта типа омега-квадрат

$$\xi_\alpha = \int_X f_\alpha(x,\omega)dF_\alpha(x,\omega) = A \Rightarrow \int_0^1 \xi^2(t)dt \,,$$

т.е. предельным распределением этой статистики является классическое распределение омега-квадрат [8], найденное как предельное для одновыборочной статистики критерия согласия омега-квадрат Крамера – Мизеса – Смирнова [1].

Действительно, сформулированное утверждение справедливо. Однако доказательство нетривиально.

Так, может показаться очевидным следующее утверждение.

Утверждение 1. Пусть f: $[0; 1] \to R^1$ – ограниченная функция, $G_n(x)$ и $G(x)$ – функции распределения, $G_n(0) = G(0) = 0$, $G_n(1) = G(1) = 1$, причем $G_n(x) \to G(x)$ при всех x. Тогда

$$\lim_{n\to\infty}\int_0^1 f(x)d(G_n(x)-G(x)) = 0\,. \qquad (4)$$

Это утверждение неверно (ср. [9, с.42]). Действительно, пусть $f(x)$ = 1, если x рационально, и $f(x) = 0$, если x иррационально, $G(x) = x$, а $G_n(x)$ имеет скачки величиной 2^{-n} в точках $m/2^n$, $m = 1, 2, \dots, 2^n$ при всех $n = 1, 2, \dots$ Тогда $G_n(x) \to G(x)$ при всех x, однако

$$\int_0^1 f(x)dG_n(x) = 1,\ \int_0^1 f(x)dG(x) = 0$$

при всех $n = 1, 2, \dots$ Следовательно, вопреки сформулированному выше утверждению 1,

$$\int_0^1 f(x)d(G_n(x)-G(x))=1,$$

т.е. соотношение (4) неверно.

Итак, сформулируем проблему. Пусть известно, что последовательность случайных функций $f_\alpha(x,\omega)$ сходится по распределению при $\alpha\to\infty$ к случайной функции $f(x,\omega)$. Пусть последовательность случайных мер $F_\alpha(A,\omega)$ сходится по распределению к вероятностной мере $F(A)$ при $\alpha\to\infty$. Если речь идет о конечномерном пространстве и меры задаются функциями распределения, то сходимость $F_\alpha(x,\omega)$ к $F(x)$ должна иметь место во всех точках непрерывности $F(x)$. В каких случаях можно утверждать, что при $\alpha\to\infty$ справедлив предельный переход

$$\xi_\alpha=\xi(f_\alpha,F_\alpha)=\int_X f_\alpha(x,\omega)dF_\alpha(x,\omega)\Rightarrow\xi=\xi(f,F)=\int_X f(x,\omega)dF(x)\,?$$

Выше показано, что, например, ограниченности $f_\alpha(x,\omega)$ для этого недостаточно.

3. Метод аппроксимации ступенчатыми функциями

Пусть $T = \{C_1, C_2, \dots , C_k\}$ – разбиение пространства X на непересекающиеся подмножества. Пусть в каждом элементе C_j разбиения T выделена точка x_j, $j = 1, 2, \dots , k$. На множестве функций f: $X\to Y$ введем оператор A_T следующим образом: если $x\in C_j$, то

$$A_Tf(x) = f(x_j),\ j = 1, 2, \dots . k. \qquad (5)$$

Тогда A_Tf – аппроксимация функции f ступенчатыми (кусочно-постоянными) функциями.

Пусть $f_\alpha(x,\omega)$ – последовательность случайных функций на X, а $K(\cdot)$ – функционал на множестве всех возможных их траекторий как функций от x. Для изучения распределения $K(f_\alpha)$ методом аппроксимации ступенчатыми функциями используют разложение

$$K(f_\alpha) = K(A_T f_\alpha) + \{K(f_\alpha) - K(A_T f_\alpha)\}. \qquad (6)$$

Согласно (5) распределение первого слагаемого в (6) определяется конечномерным распределением случайного элемента, а именно, распределением вектора

$$(f_\alpha(x_1,\omega), f_\alpha(x_2,\omega), \ldots \; f_\alpha(x_k,\omega)). \qquad (7)$$

В обычных постановках предельной теории непараметрических критериев распределение вектора (7) сходится при $\alpha\to\infty$ к соответствующему конечномерному распределению предельной случайной функции $f(x,\omega)$, т.е. к распределению случайного вектора

$$(f(x_1,\omega), f(x_2,\omega), \ldots \; f(x_k,\omega)). \qquad (8)$$

В соответствии с теорией наследования сходимости (см. [7, разд. П3]) при слабых условиях на функционал $K(\cdot)$ из сходимости по распределению вектора (7) к вектору (8) следует сходимость по распределению $K(A_T f_\alpha)$ к $K(A_T f)$.

Используя аналогичное (6) разложение

$$K(f) = K(A_T f) + \{K(f) - K(A_T f)\}, \qquad (9)$$

можно устанавливать сходимость по распределению $K(f_\alpha)$ к $K(f)$ при $\alpha\to\infty$ в два этапа: сначала выбрать разбиение T так, чтобы вторые слагаемые в правых частях соотношений (6) и (9) были малы, а затем при фиксированном операторе A_T воспользоваться сходимостью по распределению $K(A_T f_\alpha)$ к $K(A_T f)$.

Рассмотрим простой пример применения метода аппроксимации ступенчатыми функциями.

4. Обобщение теоремы Хелли

Пусть f: $[0;\ 1] \to R^1$ – измеримая функция, $F_n(x)$ – функции распределений, сосредоточенных на отрезке [0; 1]. Пусть $F_n(x)$ сходятся в основном к функции распределения $F(x)$, т.е.

$$\lim_{n\to\infty} F_n(x) = F(x) \qquad (10)$$

для всех x, являющихся точками непрерывности $F(x)$.

Утверждение 2. Если $f(x)$ – непрерывная функция, то

$$\lim_{n\to\infty}\int_0^1 f(x)dF_n(x)=\int_0^1 f(x)dF(x) \qquad (11)$$

(рассматриваются интегралы Лебега-Стилтьеса).

Утверждение 2 известно в литературе как первая теорема Хелли [3, с.344-346], вторая теорема Хелли [10, с.174-175], лемма Хелли - Брея [11, с.193-194].

Естественно поставить вопрос: при каких f из (10) следует (11)? Необходимо ввести условия и на F_n. Отметим, что если $F_n \equiv F$, то соотношение (11) верно для любой измеримой функции f, для которой интеграл в (11) существует. Поэтому рассмотрим следующую постановку.

Постановка 1. Пусть функция f такова, что для *любой* последовательности F_n, удовлетворяющей (10), справедливо (11). Что можно сказать о функции f?

В работах [4, 5] найдены следующие необходимые и достаточные условия на функцию f.

Теорема 1. Пусть ограниченная на [0; 1] функция f интегрируема по Риману-Стилтьесу по функции распределения $F(x)$. Тогда для *любой* последовательности функций распределения F_n, сходящейся в основном к F, имеет место предельный переход (11).

Теорема 2. Пусть функция f не интегрируема по Риману-Стилтьесу по функции распределения $F(x)$. Тогда *существует* последовательность функций распределения F_n, сходящаяся в основном к F, для которой соотношение (11) не выполнено.

Теоремы 1 и 2 в совокупности дают необходимые и достаточные условия для f в постановке 1. А именно, необходимо и достаточно, чтобы ограниченная на [0; 1] функция f была интегрируема по Риману-Стилтьесу по F.

Напомним определение интегрируемости функции f по Риману-Стилтьесу по функции распределения F согласно [3, с.341]. Рассмотрим разбиение $T = \{C_1, C_2, \dots , C_k\}$, где

$$C_i = [y_{i-1}, y_i),\ i = 1, 2, \dots, m-1,\ C_m = [y_{m-1}, y_m], \qquad (12)$$
$$0 = y_0 < y_1 < y_2 <\dots< y_m = 1.$$

Выберем в C_i произвольную точку x_i, $i = 1, 2, \dots, m$, и составим сумму

$$S(T) = \sum_{i=1}^{m} f(x_i)[F(y_i) - F(y_{i-1})].$$

Если при $\max(y_i - y_{i-1}) \to 0$ эти суммы стремятся к некоторому пределу (не зависящему ни от способа дробления отрезка [0; 1], ни от выбора точек x_i в каждом из элементов разбиения), то этот предел называется интегралом Римана-Стилтьеса от функции f по функции F по отрезку [0; 1] и обозначается символом, приведенным в правой части равенства (11).

Рассмотрим суммы Дарбу-Стилтьеса

$$S_H(T) = \sum_{i=1}^{m} m_i[F(y_i) - F(y_{i-1})], \quad S_B(T) = \sum_{i=1}^{m} M_i[F(y_i) - F(y_{i-1})],$$

где

$$m_i = \inf\{ f(x), x \in X_i\}, \quad M_i = \sup f\{ f(x), x \in X_i\}.$$

Ясно, что

$$S_H(T) \le S(T) \le S_B(T).$$

Необходимым и достаточным условием интегрируемости по Риману-Стилтьесу является следующее: для любой последовательности разбиений T_k, $k = 1, 2, 3, \dots$ вида (12) такой, что $\max(y_i - y_{i-1}) \to 0$ при $k\to\infty$, имеем

$$\lim_{k\to\infty}[S_B(T_k) - S_H(T_k)] = 0. \qquad (13)$$

Напомним, что колебанием $\delta(f, B)$ функции f на множестве B называется $\delta(f, B) = \sup\{|f(x) - f(y)|, x\in B, y\in B\}$. Поскольку

$$\delta(f, C_i) = M_i - m_i,$$

то условие (13) можно записать в виде

$$\lim_{k\to\infty}\sum_{C\in T_k}\delta(f,C)F(C)=0\,. \qquad (14)$$

Условие (14), допускающее обобщение с X = [0; 1] и f: [0; 1] → R^1 на X и f более общего вида, и будем использовать при доказательстве теорем 1 и 2.

Доказательство теоремы 1. Согласно методу аппроксимации ступенчатыми функциями рассмотрим оператор A_T. Как легко проверить, имеет место разложение

$$\begin{aligned}\beta_n=\int_0^1 f(x)dF_n(x)-\int_0^1 f(x)dF(x)=\int_0^1\{f(x)-A_T f(x)\}dF_n(x)+\\+\int_0^1\{A_T f(x)-f(x)\}dF(x)+\left\{\int_0^1 A_T f(x)dF_n(x)-\int_0^1 A_T f(x)dF(x)\right\}\end{aligned}. \qquad (15)$$

Поскольку

$$|f(x) - A_T f(x)| \le \delta(f, X_i),\ \ x\in C_i,$$

то первое слагаемое в правой части (15) не превосходит

$$\sum_{C\in T}\delta(f,C)F_n(C)\,, \qquad (16)$$

а второе не превосходит

$$\sum_{C\in T}\delta(f,C)F(C)\,.$$

Согласно определению оператора A_T третье слагаемое в (15) имеет вид

$$\sum_{i=1}^{m} f(x_i)(F_n(C_i)-F(C_i))\,.$$

Очевидно, оно не превосходит по модулю

$$\sup_{x\in X}|f(x)|\sum_{C\in T}|F_n(C)-F(C)|$$

(здесь используется ограниченность f на X).

Согласно (16) первое слагаемое в правой части (15) не превосходит

$$\sum_{C\in T}\delta(f,C)F(C)+\sum_{C\in T}\delta(f,C)\,|F_n(C)-F(C)|\,.$$

Поскольку

$$\delta(f,C)\le 2\sup_{x\in X}|f(x)|\,,$$

то первое слагаемое в правой части (15) не превосходит

$$\sum_{C\in T}\delta(f,C)F(C)+2\sup_{x\in X}|f(x)|\sum_{C\in T}|F_n(C)-F(C)|.$$

Из оценок, относящихся к трем слагаемым в разложении (15), следует, что

$$|\beta_n|\leq 2\sum_{C\in T}\delta(f,C)F(C)+3\sup_{x\in X}|f(x)|\sum_{C\in T}|F_n(C)-F(C)|. \qquad (17)$$

Используя оценку (17), докажем, что $\beta_n \to 0$ при $n \to \infty$. Пусть дано ε > 0. Согласно условию интегрируемости функции f по Риману-Стилтьесу, т.е. условию (14), можно указать разбиение $T = T(\varepsilon)$ такое, что

$$\sum_{C\in T(\varepsilon)}\delta(f,C)F(C)<\frac{\varepsilon}{4}, \qquad (18)$$

и в точках y_i, $i = 1, 2, \ldots, m - 1$ (см. (12)), функция F непрерывна.

Поскольку

$$F_n(X_i) = F_n(y_i) - F_n(y_{i-1}),$$

то из (10) следует, что существует число $n = n(\varepsilon)$ такое, что при $n > n(\varepsilon)$ справедливо неравенство

$$\sum_{C\in T(\varepsilon)}|F_n(C)-F(C)|<\frac{\varepsilon}{6}\left(\sup_{x\in X}|f(x)|\right)^{-1}. \qquad (19)$$

Из (17), (18) и (19) следует, что при $n > n(\varepsilon)$ справедливо неравенство

$$\left|\int_0^1 f(x)dF_n(x)-\int_0^1 f(x)dF(x)\right|<\varepsilon,$$

что и требовалось доказать.

Обсудим условие ограниченности f. Если оно не выполнено, то из (10) не всегда следует (11).

Пример 2. Пусть $f(x) = 1/x$ при $x > 0$ и $f(0)=0$. Пусть $F(0,5) = 0$, т.е. предельное распределение сосредоточено на [1/2; 1]. Пусть распределение F_n на [0; S) имеет единственный атом в точке $x = 1/n$ величиной $n^{-1/2}$, а на [1/2; 1] справедливо (10). Тогда по причинам, изложенным при доказательстве теоремы 1,

$$\lim_{n\to\infty}\int_{1/2}^{1} f(x)dF_n(x)=\int_{1/2}^{1} f(x)dF(x),$$

однако

$$\int_{0}^{1/2} f(x)dF_n(x)=\sqrt{n},\quad \int_{0}^{1/2} f(x)dF(x)=0,$$

т.е. соотношение (11) не выполнено.

Условие ограниченности подынтегральной функции f можно заменить, как это сделано, например, в [4], на условие строгого возрастания функции распределения F.

Лемма. Пусть функции распределения F всюду строго возрастает, т.е. из $x_1 < x_2$ вытекает $F(x_1) < F(x_2)$. Пусть функция f интегрируема по Риману-Стилтьесу по F, т.е. выполнено (14). Тогда функция f ограничена.

Доказательство. Рассмотрим точки $0 = y_0 < y_1 < y_2 <\ldots< y_{2m} = 1$ и два разбиения

$$T_1=\{[0;y_1),[y_1;y_3),[y_3;y_5),...,[y_{2m-1};1]\},\ T_2=\{[0;y_2),[y_2;y_4),[y_4;y_6),...,[y_{2m-2};1]\}.$$

Тогда для любых двух точек x и x' можно указать конечную последовательность точек $x_1 = x$, x_2, x_3, …, x_s, $x_{s+1} = x'$ такую, что любые две соседние точки x_i, x_{i+1}, i = 1, 2, …, s, одновременно принадлежат некоторому элементу C_i разбиения T_1 или разбиения T_2, причем $C_i \neq C_j$ при $i \neq j$. Действительно, пусть $x\in[y_p;y_{p+1}),x'\in[y_q;y_{q+1})$. Пусть для определенности $q > p$. Тогда можно положить $x_2 = y_{p+1}$, $x_3 = y_{p+2}$, …, $x_s = y_q$. Поскольку среди элементов разбиений T_1 и T_2 есть $C_1 = [y_p;\ y_{p+2})$, то $x\in x_1\in C_1, x_2=y_{p+1}\in C_1$. Далее, $x_2\in[y_{p+1},y_{p+3})=C_2, x_3\in C_2$, и т.д.

Из указанных выше свойств последовательности $x_1 = x$, x_2, x_3, …, x_s, $x_{s+1} = x'$ следует, что

$$|f(x)-f(x')|\le\sum_{i=1}^{s}|f(x_{i+1})-f(x_i)|\le\sum_{C\in T_1}\delta(f,C)+\sum_{C\in T_2}\delta(f,C).$$

Пусть теперь число $\max(y_i - y_{i-2})$ настолько мало, что согласно (14)

$$\sum_{C\in T_1}\delta(f,C)F(C)<1,\ \sum_{C\in T_2}\delta(f,C)F(C)<1.$$

Тогда согласно двум последним соотношениям

$$|f(x)-f(x')|\le 2[\min\{F(C):C\in T_1\bigcup T_2\}]^{-1},$$

что и доказывает лемму.

Доказательство теоремы 2. Пусть условие (14) не выполнено, т.е. существуют число $\gamma > 0$ и последовательность разбиений T_n, $n = 1, 2, \ldots$, такие, что $\max(y_i - y_{i-1}) \to 0$ при $n\to\infty$ и при всех n

$$\sum_{C\in T_n}\delta(f,C)F(C)\ge\gamma. \qquad (20)$$

Для доказательства теоремы построим две последовательности функций распределения F_{1n} и F_{2n}, $n = 1, 2, \ldots$, для которых выполнено (10), но последовательность

$$\delta_n=\int_0^1 f(x)dF_{1n}(x)-\int_0^1 f(x)dF_{2n}(x)$$

не стремится к 0 при $n \to \infty$. Тогда (11) не выполнено хотя бы для одной из последовательностей F_{1n} и F_{2n}.

Для любого C – элемента некоторого разбиения T – можно указать, как вытекает из определения $\delta(f, C)$, точки $x_1(C)$ и $x_2(C)$ такие, что

$$f(x_1(C)) - f(x_2(C)) > S\ \delta(f, C). \qquad (21)$$

Построим F_{1n} и F_{2n} следующим образом. Пусть $F_{1n}(C) = F_{2n}(C) = F(C)$ для любого C из T_n. При этом F_{1n} имеет в C один атом в точке $x_1(C)$ величиной $F(C)$, а F_{2n} имеет в C также один атом в точке $x_2(C)$ той же величины $F(C)$. Другими словами, распределение F_{1n} в C сосредоточено в одной точке, а именно, в $x_1(C)$, а распределение F_{2n} сосредоточено в $x_2(C)$. Тогда

$$\delta_n=\sum_{C\in T_n}(f(x_1(C))-f(x_2(C)))F(C). \qquad (22)$$

Из (20), (21) и (22) следует, что

$$\delta_n \ge \frac{1}{2}\sum_{C\in T_n}\delta(f,C)F(C) \ge \frac{\gamma}{2}.$$

Остается показать, что для последовательностей функций распределения F_{1n} и F_{2n} выполнено (10). Пусть x – точка непрерывности F. Пусть

$$y_1(x, T) = \max\{y_{kn}: y_{kn} < x\},\ y_2(x, T) = \min\{ y_{kn}: y_{kn} > x\},$$

где y_{kn} – точки, определяющие разбиения T_n согласно (12). В соответствии с определением F_{in}

$$F_{in}(y_j(x, T_n))= F(y_j(x, T_n)),\ i = 1, 2,\ j = 1, 2,$$

а потому

$$|F_{in}(\mathrm{x}) – F(x)| \le F(y_2(x, T_n)) - F(y_1(x, T_n)),\ i = 1, 2.$$

В силу условия $\max(y_{kn} – y_{(k-1)n}) \to 0$ и непрерывности F в точке x правая часть последнего соотношения стремится к 0 при $n \to \infty$, что и заканчивает доказательство теоремы 2.

Теоремы 1 и 2 демонстрируют основные идеи предельной теории статистик интегрального типа и непараметрических критериев в целом. Как показывают эти теоремы, основную роль в рассматриваемой теории играет предельное соотношение (14). Отметим, что если $\delta(f, T_n) \to 0$ при $n \to \infty$, то (14) справедливо, но, вообще говоря, не наоборот. Естественно возникает еще ряд постановок. Пусть (14) выполнено для f_1 и f_2. При каких функциях h это соотношение выполнено для $h(x, f_1(x), f_2(x))$? В прикладной статистике вместо $f(x)$ рассматривают $f_\alpha(x, \omega)$ и $f(x, \omega)$, а вместо интегрирования по функциям распределения $F_n(x)$ – интегрирование по случайным мерам $F_\alpha(\omega)$. Как меняются формулировки в связи с такой заменой? В связи со слабой сходимостью (т.е. сходимостью по распределению) A_Tf_α к A_T и переходом от $f_\alpha(x, \omega)$ к $h_\alpha(x, f_{1\alpha}(x, \omega), f_{2\alpha}(x, \omega))$ возникает следующая постановка. Пусть последовательность случайных элементов κ_α слабо сходится к случайному элементу к при $\alpha\to\infty$. Когда распределения случайных элементов $g_\alpha(\kappa_\alpha)$ сближаются с распределениями случайных элементов g_α(к)? Полным ответом на последний вопрос являются необходимые и достаточные условия наследования сходимости. Они приведены в [7, разд.П3].

5. Основные результаты о статистиках интегрального типа

Наиболее общая теорема типа теоремы 1 выглядит так [5].

Теорема 3. Пусть существует последовательность разбиений T_n, n = 1, 2, …, такая, что при $n \to\infty$ и $\alpha\to\infty$

$$\Delta(f_\alpha, T_n) = \sum_{C \in T_n} \delta(f_\alpha, C) F(C) \to 0\,. \qquad (23)$$

Пусть для любого C, входящего хотя бы в одно из разбиений T_n,

$$F_\alpha(C, \omega) \to F(C) \qquad (24)$$

при $\alpha\to\infty$ (сходимость по вероятности). Пусть f_α асимптотически ограничены по вероятности при $\alpha\to\infty$. Тогда

$$\xi(f_\alpha, F_\alpha) - \xi(f_\alpha, F) \to 0 \qquad (25)$$

при $\alpha\to\infty$ (сходимость по вероятности).

Как известно, полное сепарабельное метрическое пространство называется польским. Это понятие понадобится для формулировки аналога теоремы 2.

Теорема 4. Пусть X – польское пространство, У конечномерно, существует измельчающаяся последовательность T_n разбиений, для которой соотношение (23) не выполнено. Тогда существует удовлетворяющая (24) последовательность F_α, для которой соотношение (25) неверно, хотя F_α слабо сходится к F при $\alpha\to\infty$.

Условие (23) естественно назвать условием римановости, поскольку в случае, рассмотренном в теореме 1, оно является условием интегрируемости по Риману-Стилтьесу. Рассмотрим *наследуемость римановости* при переходе от $f_{1\alpha}(x, \omega)$ со значениями в $У_1$ и $f_{2\alpha}(x, \omega)$ со значениями в $У_2$, удовлетворяющих (23), к $h_\alpha(x,\ f_{1\alpha}(x,\ \omega),\ f_{2\alpha}(x,\ \omega))$ со значениями в $У_3$.

Положим

$$Y_k(a,\varepsilon)=\{(y,y'): y\in Y_k, y'\in Y_k, \|y\|_k<a, \|y'\|<a, \|y-y'\|<\varepsilon\}, k=1,2,$$

где $\|\cdot\|_k$ – норма (т.е. длина вектора) в пространстве Y_k, k = 1, 2. Рассмотрим также множества

$$A(C,a,\varepsilon)=\{(x,x',y_1,y_1^*,y_2,y_2^*): x,x'\in C, (y_k,y_k^*)\in Y_k(a,\varepsilon), k=1,2\}$$

и функции

$$q_\alpha(x,x',y_1,y_1^*,y_2,y_2^*)=h_\alpha(x,y_1,y_2)-h_\alpha(x',y_1^*,y_2^*).$$

Наконец, понадобится измеритель колеблемости

$$c(h_\alpha,T,a,\varepsilon)=\sum_{C\in T}\sup_{A(C,a,\varepsilon)}\|q_\alpha\|_3\, F(C)$$

и множество

$$Z(a)=X\times\{y_1:\|y_1\|<a\}\times\{y_2:\|y_2\|<a\}.$$

Теорема 5. Пусть h_α асимптотически (при $\alpha\to\infty$) ограничены на $Z(a)$ при любом положительном a, функции $f_{1\alpha}$ и $f_{2\alpha}$ асимптотически ограничены по вероятности и удовлетворяют условию (23). Пусть для участвующей в (23) последовательности T_n

$$c(h_\alpha,\ T_n,\ a,\ \varepsilon)\to 0 \qquad (26)$$

при $\alpha\to\infty$, $n\to\infty$, $\varepsilon\to 0$ и любом положительном a. Тогда $f_{3\alpha}(x,\ \omega) = h_\alpha(x,\ f_{1\alpha}(x,\ \omega),\ f_{2\alpha}(x,\ \omega))$ удовлетворяют условию (23) и асимптотически ограничены по вероятности.

Теорема 6. Пусть условие (26) не выполнено для h_α. Тогда существуют детерминированные ограниченные функции $f_{1\alpha}$ и $f_{2\alpha}$ такие, что соотношение (23) выполнено для $f_{1\alpha}$ и $f_{2\alpha}$ и не выполнено для $f_{3\alpha}$.

Пример 3. Пусть $X = [0;\ 1]^k$, пространства Y_1 и Y_2 конечномерны, функция $h_\alpha \equiv h(x,\ y_1,\ y_2)$ непрерывна. Тогда условие (26) выполнено.

С помощью теорем 3 и 5 и результатов о наследовании сходимости можно изучить асимптотическое поведение статистик интегрального типа

$$\xi_\alpha = \int_X h_\alpha(x, f_{1\alpha}(x,\omega), f_{2\alpha}(x,\omega)) F_\alpha(dx,\omega)$$

со значениями в банаховом пространстве *У*.

Теорема 7. Пусть для некоторой последовательности T_n разбиений X справедливы соотношения (23) для $f_{1\alpha}$ и $f_{2\alpha}$ и (24) для F_α. Пусть последовательность функций h_α удовлетворяет условию в теореме 5, конечномерные распределения $(f_{1\alpha}(x, \omega), f_{2\alpha}(x, \omega))$ слабо сходятся к конечномерным распределениям $(f_1(x, \omega), f_2(x, \omega))$, причем для f_1 и f_2 справедливо соотношение (23). Тогда

$$\lim_{\alpha\to\infty} L(\xi_\alpha, \eta_\alpha) = 0,$$

где L – расстояние Прохорова (см. [7, разд. П.3]),

$$\eta_\alpha = \int_X h_\alpha(x, f_1(x,\omega), f_2(x,\omega)) F(dx).$$

Теорема 7 дает общий метод получения асимптотических распределений статистик интегрального типа. Важно, что соотношение (23) выполнено для эмпирического процесса и для процессов, связанных с оцениванием параметров при проверке согласия [1, 4].

Один из выводов общей теории состоит в том, что в качестве F_α можно использовать практически любую состоятельную оценку истинной функции распределения. Этот вывод использовался при построении критерия типа омега-квадрат для проверки симметрии распределения относительно 0 и обнаружения различий в связанных выборках [12, 13]).

6. Критерии типа Колмогорова - Смирнова

Асимптотическое поведение критериев типа Колмогорова – Смирнова (их называют также критериями супремумного типа) может быть получено с помощью описанного выше метода аппроксимации ступенчатыми функциями. Этот метод не требует обращения к теории сходимости вероятностных мер в функциональных пространствах [14].

Для критериев Колмогорова и Смирнова достаточно использовать лишь свойства эмпирического процесса и броуновского моста. В случае проверки согласия с помощью критерия типа Колмогорова – Смирнова [1] добавляется необходимость изучения еще одного случайного процесса. Он является разностью между двумя функциями распределения. Одна – (теоретическая) функция распределения элементов выборки. Вторая - член параметрического семейства распределений, полученный путем подстановки оценок параметров вместо их истинных значений.

7. Выводы

Итак, асимптотическое поведение статистик интегрального и супремумного типов может быть изучено с помощью полученных в настоящей статье результатов, основанных на методе аппроксимации ступенчатыми функциями. Предельные распределения практически используемых статистик рассматриваемых типов находят как распределения функционалов от соответствующих случайных процессов (в частности, броуновских процессов, броуновских мостов). Для статистик типа омега-квадрат математических аппарат расчета предельных распределений развит в [9]. Важно, что настоящая статья дает полное обоснование эвристическому подходу Дуба к получению предельных теорем и показывает тем самым, что для изучения асимптотики практически используемых статистик интегрального и супремумного типов (т.е. статистик типов Колмогорова – Смирнова и омега-квадрат) нет необходимости привлекать теорию сходимости вероятностных мер в функциональных пространствах.

В нескольких постановках найдены необходимые и достаточные условия. Это означает, что полученные результаты являются окончательными.

Литература

1. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении / А.И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета (Научный журнал КубГАУ) [Электронный ресурс]. – Краснодар: КубГАУ, 2014. – №03(097). С. 647 – 675. – IDA [article ID]: 0971403047. – Режим доступа: http://ej.kubagro.ru/2014/03/pdf/47.pdf
2. Келли Дж. Общая топология. М.: Наука, 1968. 384 с.
3. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. М.: Наука, 1972. 496 с.
4. Орлов А.И. Асимптотическое поведение статистик интегрального типа // Доклады АН СССР. 1974. Т.219. № 4. С. 808-811.
5. Орлов А.И. Асимптотическое поведение статистик интегрального типа // Вероятностные процессы и их приложения. Межвузовский сборник научных трудов. М.: МИЭМ, 1989. С.118-123.
6. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч.3. Статистические методы анализа данных. М.: Изд-во МГТУ им. Н.Э. Баумана, 2012. 624 с.
7. Орлов А.И. Организационно-экономическое моделирование : учебник : в 3 ч. Ч. 1. Нечисловая статистика. М.: Изд-во МГТУ им. Н.Э. Баумана, 2009. 541 с.
8. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983. 416 с.
9. Мартынов Г.В. Критерии омега-квадрат. М.: Наука, 1978. 80 с.
10. Гнеденко Б.В. Курс теории вероятностей: 7-е изд., исправл. М.: Эдиториал УРСС, 2001. 320 с.
11. Лоэв М. Теория вероятностей. М.: ИЛ, 1962. 720 с.
12. Орлов А.И. О проверке симметрии распределения // Теория вероятностей и ее применения. 1972. Т.17. № 2. С.372-377.
13. Орлов А.И. Методы проверки однородности связанных выборок // Заводская лаборатория. Диагностика материалов. 2004. Т.70. № 7. С.57-61.
14. Биллингсли П. Сходимость вероятностных мер. М.: Наука, 1977. 352 с.

References

1. Orlov A.I. Neparametricheskie kriterii soglasija Kolmogorova, Smirnova, omega-kvadrat i oshibki pri ih primenenii / A.I. Orlov // Politematicheskij setevoj jelektronnyj nauchnyj zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta (Nauchnyj zhurnal KubGAU) [Jelektronnyj resurs]. – Krasnodar: KubGAU, 2014. – №03(097). S. 647 – 675. – IDA [article ID]: 0971403047. – Rezhim dostupa: http://ej.kubagro.ru/2014/03/pdf/47.pdf
2. Kelli Dzh. Obshhaja topologija. M.: Nauka, 1968. 384 s.
3. Kolmogorov A.N., Fomin S.V. Jelementy teorii funkcij i funkcional'nogo analiza. M.: Nauka, 1972. 496 s.
4. Orlov A.I. Asimptoticheskoe povedenie statistik integral'nogo tipa // Doklady AN SSSR. 1974. T.219. № 4. S. 808-811.
5. Orlov A.I. Asimptoticheskoe povedenie statistik integral'nogo tipa // Verojatnostnye processy i ih prilozhenija. Mezhvuzovskij sbornik nauchnyh trudov. M.: MIJeM, 1989. S.118-123.
6. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie : uchebnik : v 3 ch. Ch.3. Statisticheskie metody analiza dannyh. M.: Izd-vo MGTU im. N.Je. Baumana, 2012. 624 s.

7. Orlov A.I. Organizacionno-jekonomicheskoe modelirovanie : uchebnik : v 3 ch. Ch. 1. Nechislovaja statistika. M.: Izd-vo MGTU im. N.Je. Baumana, 2009. 541 s.
8. Bol'shev L.N., Smirnov N.V. Tablicy matematicheskoj statistiki. M.: Nauka, 1983. 416 s.
9. Martynov G.V. Kriterii omega-kvadrat. M.: Nauka, 1978. 80 s.
10. Gnedenko B.V. Kurs teorii verojatnostej: 7-e izd., ispravl. M.: Jeditorial URSS, 2001. 320 s.
11. Lojev M. Teorija verojatnostej. M.: IL, 1962. 720 s.
12. Orlov A.I. O proverke simmetrii raspredelenija // Teorija verojatnostej i ee primenenija. 1972. T.17. № 2. S.372-377.
13. Orlov A.I. Metody proverki odnorodnosti svjazannyh vyborok // Zavodskaja laboratorija. Diagnostika materialov. 2004. T.70. № 7. S.57-61.
14. Billingsli P. Shodimost' verojatnostnyh mer. M.: Nauka, 1977. 352 s.