1953 УКРАИНСКИЙ МАТЕМАТИЧЕСКИЙ ЖУРНАЛ Т. V, № 4
ИНСТИТУТ МАТЕМАТИКИ

О некоторых предельных теоремах для условных распределений и о связанных с ними задачах математической статистики

И. И. Гихман

Настоящая статья состоит из двух частей. Хотя задачи, рассматриваемые в них, на первый взгляд, кажутся мало связанными, в действительности же их объединяет общность метода исследования и внутреннее единство. В первой части доказываются некоторые теоремы, относящиеся к условному распределению функционалов от последовательности независимых случайных величин, во второй — рассматриваются вопросы, связанные с распределением колмогоровского критерия согласия в том случае, когда проверяемая функция распределения содержит параметры, определяемые эмпирическим путем.

I

Ряд задач теории вероятностей и математической статистики, рассмотренных в последнее время, являются частными случаями следующей более общей задачи.

Дана последовательность серий

$$\Gamma_n : \xi_{n1}, \xi_{n2}, \ldots, \xi_{n, N_n} \tag{1}$$

независимых одинаково распределенных случайных величин и некоторое число измеримых функционалов $f_k(\Gamma_n)$, $k = 1, 2, \ldots, m$, от последовательностей Γ_n. Требуется определить совместное условное распределение случайных величин $f_k(\Gamma_n)$ при гипотезе $\sum_{r=1}^{N_n} \xi_{nr} = z_n$.

В большинстве случаев речь идет об асимптотическом распределении функционалов, когда $n \to \infty$ и серии Γ_n удовлетворяют определенным условиям.

Представляет интерес также и тот случай, когда число членов в последовательности Γ_n само является случайной величиной. Последнее будет, например, в случае, когда Γ_n есть последовательность скачков в течение определенного промежутка времени случайной функции некоторого однородного разрывного процесса.

Вот несколько примеров функционалов, представляющих интерес для математической статистики:

$$M_n(t)=\max_{0<k<tN_n}\sum_{r=1}^{k}\xi_{nr}, \qquad \sum_{1}^{0}\xi_{nr}=0, \quad 0<t\leqslant 1; \qquad m_n(t)=-\min_{0<k<tN_n}\sum_{r=1}^{k}\xi_{nr};$$

$I_n(h, t)$ — число сумм $\eta_{nk}=\sum_{r=1}^{k}\xi_{nr}$, $0\leqslant k\leqslant tN_n$, превосходящих число $h\geqslant 0$;

$I_n(h_1, h_2; t)$ — число сумм η_{nk}, где $0\leqslant k\leqslant tN_n$, значения которых лежат вне интервала $(-h_1, h_2)$.

Приведем теперь примеры конкретных задач.

1. Случайные величины ξ_k принимают значения ± 1 с вероятностью $\frac{1}{2}$, $\xi_{nk}=\frac{1}{\sqrt{n}}\xi_k$, $N_n=n$. Распределение функционала $I_n(0, 1)$ при конечном n было рассмотрено в работах [1], [2] (и в несколько более общем случае в [3]).

Предельная теорема для $I_n(0, 1)$, $n\to\infty$ была дана в [4] для случая, когда ξ_k принимают целые значения.

При конечном n совместное распределение величин $m_n(1)$, $M_n(1)$ было найдено в [5]. Задолго до этого предельная теорема для случайных величин, принимающих только два значения, была установлена Н. В. Смирновым [6].

Отметим, что в работах Н. В. Смирнова, Б. В. Гнеденко и его учеников соответствующие задачи формулировались несколько иначе, непосредственно как задачи математической статистики, относящиеся к проверке однородности двух эмпирических функций распределения.

2. Пусть $\zeta(t, N)$ — случайная функция пуассоновского процесса. Это означает, что $\zeta(t, N)$ меняется скачками, величина которых равна 1, причем среднее число скачков за промежуток времени Δt равно $N\,\Delta t+o(\Delta t)$ и количества скачков в течение непересекающихся интервалов времени независимы.

Случайная функция

$$\frac{1}{N}\zeta(t, N)-t, \quad 0\leqslant t\leqslant 1,$$

при гипотезе $\zeta(1, N)=N$ эквивалентна функции

$$\sqrt{N}\{F_N(x)-F(x)\}, \quad t=F(x), \tag{2}$$

где $F_N(x)$ — эмпирическая функция распределения, построенная по N независимым измерениям некоторой случайной величины с непрерывной функцией распределения $F(x)$. Распределение абсолютного максимума функции (2) было предметом известной работы А. Н. Колмогорова [7], явившейся основополагающей для многочисленных дальнейших исследований.

3. Пусть $\{\xi_k\}$ — одинаково распределенные независимые величины и

$$\sqrt{n}\,\sigma^{\bullet}_{nk}=\sum_{1}^{k}\xi_r-\frac{k}{n}\sum_{1}^{n}\xi_r. \tag{3}$$

Предельное безусловное распределение максимального и минимального члена последовательности σ^*_{nk} без доказательства было указано Феллером [8]. В дальнейшем окажется, что предельное условное распределение экстремальных членов последовательности (3) не зависит от гипотезы z, $\sum_1^n \xi_r = \sqrt{n} z_n$, $z_n \to z$.

4. Отметим пока только, не вдаваясь в подробности, что к задаче на определение условных распределений максимальных и минимальных членов последовательности случайных величин можно прийти рассматривая предельное распределение критерия согласия Колмогорова для того случая, когда проверяемое распределение содержит параметры, определяемые по эмпирическим данным. Этому вопросу посвящена вторая часть работы.

Первая существенная трудность, возникающая при решении задач на отыскание условных вероятностей, связана со сложностью самого понятия условной вероятности. Чтобы избежать затруднений, возникающих в общем случае, мы в дальнейшем ограничимся рассмотрением случайных величин двух типов: а) решетчатых, б) имеющих плотность распределения. Последнее требование можно заменить другим, обеспечивающим применение предельной теоремы для плотностей (по поводу предельной теоремы для плотностей распределения см., например, [9]).

Теорема 1. Допустим, что последовательность независимых одинаково распределенных случайных величин

$$\xi_1, \xi_2, \ldots, \xi_n, \ldots \quad M\xi_k = 0, \; M\xi_k^2 = \sigma^2 < \infty \tag{4}$$

удовлетворяет одному из следующих двух условий:

а) ξ_n — решетчатые величины,

б) ξ_n обладают плотностью распределения, являющейся функцией ограниченной вариации.

Пусть $\varphi(t)$, $\psi(t)$ — непрерывно дифференцируемые на интервале (0,1) функции, причем $\varphi(0) < 0 < \psi(0)$, $\varphi(t) < \psi(t)$ $(0 \leqslant t \leqslant 1)$ и z_n — одно из значений, которые может принимать величина η_{nn}, где

$$\sqrt{n}\,\sigma\eta_{nk} = \sum_1^k \xi_r.$$

Тогда при $n \to \infty$, $z_n \to z$ условная вероятность совмещения неравенств

$$\varphi\left(\frac{k}{n}\right) < \eta_{nk} < \psi\left(\frac{k}{n}\right), \quad 0 < k \leqslant Tn, \quad T < 1 \tag{5}$$

при гипотезе $\eta_{nn} = z_n$ стремится к $u(0, 0)$, где $u(t, s)$ есть решение уравнения

$$\frac{\partial u}{\partial t} + \frac{z - s}{1 - t}\frac{\partial u}{\partial s} + \frac{1}{2}\frac{\partial^2 u}{\partial s^2} = 0 \tag{6}$$

в области $\varphi(t) < s < \psi(t)$, $0 \leqslant t \leqslant T$, удовлетворяющее граничным условиям

$$u(t, \varphi(t)) = u(t, \psi(t)) = 0, \quad 0 \leqslant t < T; \qquad u(T, s) = 1, \quad \varphi(T) < s < \psi(T). \tag{7}$$

Доказательству этой теоремы мы предпошлем лемму, устанавливающую асимптотические $(n \to \infty)$ выражения для условных, „урезанных" моментов величины ξ_{k+1} при заданной сумме $n^{-\frac{1}{2}} \sum_{k+1}^{n} \xi_r$.

Лемма 1. Пусть последовательность случайных величин (4) удовлетворяет условиям теоремы 1. Обозначим через $\alpha_k(n, \varepsilon, z_n)$ условное математическое ожидание величины $\chi_\varepsilon(\xi_1)\xi_1^k$ $(k=0, 1, 2)$ при гипотезе

$$\eta_n = \frac{1}{\sqrt{n}} \sum_{1}^{n} \xi_k = z_n,$$

где z_n в случае а) — одно из допустимых значений суммы η_n, а $\chi_\varepsilon(\xi) = 1$, если $|\xi| < \varepsilon\sqrt{n}$, и $\chi_\varepsilon(\xi) = 0$, если $|\xi| \geqslant \varepsilon\sqrt{n}$.
Тогда

$$1 - \alpha_0(n, \varepsilon, z_n) = \frac{o_\varepsilon(1)}{n},$$

$$\alpha_1(n, \varepsilon, z_n) = (z_n + o_\varepsilon(1) + O(\varepsilon))\frac{1}{n},$$

$$\alpha_2(n, \varepsilon, z_n) = (1 + o_\varepsilon(1) + O(\varepsilon))\frac{1}{n},$$

где $o_\varepsilon(1) \to 0$ равномерно относительно z_n при любом фиксированном $\varepsilon > 0$ $(n \to \infty)$.

Доказательство. Ограничимся рассмотрением решетчатых случайных величин. Для случая существования плотности распределения доказательство проводится аналогично. В последнем случае вместо предельной локальной теоремы следует пользоваться предельной теоремой для плотностей.

Величины ξ_k могут принимать только значения вида

$$x_s = a + sh, \quad s = 0, \pm 1, \pm 2, \ldots,$$

где h — максимальный шаг распределения, а величины η_n значения

$$z_{ns} = \frac{h(s - n\bar{p})}{\sqrt{n}},$$

где

$$\bar{p} = -\frac{a}{h} = \sum_{-\infty}^{\infty} s p_s,$$

$$p_s = \mathrm{P}\{\xi_k = a + sh\}.$$

Пусть

$$P_n(s) = \mathrm{P}\{\eta_n = z_{ns}\}.$$

В силу локальной теоремы для решетчатых случайных величин

$$\sqrt{n} P_n(s) = \frac{h}{\sqrt{2\pi}} e^{-\frac{z_{ns}^2}{2}} + o(1), \tag{8}$$

где оценка $o(1)$ равномерна относительно z_{ns}, $-\infty < z_{ns} < \infty$.

Пусть $P_n'(s)$ — вероятность того, что величина

$$\frac{1}{\sqrt{n}}\sum_{1}^{n-1}\xi_r$$

принимает значение $\frac{h(s-\overline{n-1}\,p)}{\sqrt{n}}$, и $x_{nr}=\frac{a+rh}{\sqrt{n}}$. Тогда совместная вероятность равенств $\xi_n=a+rh$, $\eta_n=z_{ns}$ равна

$$P_n(s)\,\mathrm{P}\{x_{nr}/z_{ns}\}=p_rP_n'(s-r), \tag{9}$$

где $\mathrm{P}\{x_{nr}/z_{ns}\}$ — условная вероятность равенства $\xi_n=\sqrt{n}\,x_{nr}$ при гипотезе $\eta_n=z_{ns}$. Из (9) следует

$$\mathrm{P}\{x_{nr}/z_{ns}\}=\frac{p_rP_n'(s-r)}{P_n(s)}. \tag{10}$$

Для доказательства леммы следует оценить величины

$$\delta(\varepsilon, n, z_{ns})=\sum_{|x_{nr}|>\varepsilon}\mathrm{P}\{x_{nr}/z_{ns}\}, \tag{11}$$

$$\alpha_1(\varepsilon, n, z_{ns})=\sum_{|x_{nr}|<\varepsilon}x_{nr}\,\mathrm{P}\{x_{nr}/z_{ns}\}, \tag{12}$$

$$\alpha_2(\varepsilon, n, z_{ns})=\sum_{|x_{nr}|<\varepsilon}x_{nr}^2\,\mathrm{P}\{x_{nr}/z_{ns}\}. \tag{13}$$

Из (8) для $P_n'(s-r)$ следует

$$\sqrt{n-1}\,P_n'(s-r)=\frac{h}{\sqrt{2\pi}}e^{-\frac{(z_{ns}-x_{nr})^2}{2}}+o(1).$$

Условимся в дальнейшем рассматривать некоторый фиксированный интервал $(-A, A)$ и будем предполагать, что значения z_{ns}, x_{nr} лежат именно в этом интервале. Тогда равномерно относительно z_{ns}, x_{nr}

$$\mathrm{P}\{x_{nr}/z_{ns}\}=p_r\,e^{-\frac{x_{nr}^2}{2}+x_{nr}z_{ns}}(1+o(1)). \tag{14}$$

Чтобы оценить сумму $\delta(\varepsilon, n, z_{ns})$, заметим, что при $|z_{ns}|<A$ и при любых значениях x_{nr}

$$\left|\frac{P_n'(s-r)}{P_n(s)}\right|<C.$$

Так как

$$\sum_{-\infty}^{\infty}(a+rh)^2p_r=1,$$

то

$$\delta(\varepsilon, n, z_{ns})\leqslant C\sum_{|x_{nr}|>\varepsilon}p_r\leqslant\frac{C}{\varepsilon^2}\sum_{|x_{nr}|>\varepsilon}x_{nr}^2p_r=\frac{o_\varepsilon(1)}{n},$$

где $o_\varepsilon(1)$ — величина равномерно сходящаяся (относительно z_{ns}) к нулю при любом фиксированном $\varepsilon>0$.

Перейдем теперь к величине $\alpha_1(\varepsilon, n, z_{ns})$. Имеем

$$\alpha_1(\varepsilon, n, z_{ns}) = \sum_{|x_{nr}|<\varepsilon} x_{nr} e^{-\frac{x_{nr}^2}{2} + x_{nr} z_{ns}} p_r (1+o(1)) =$$

$$= \left\{ \sum_{|x_{nr}|<\varepsilon} x_{nr} p_r + \sum_{|x_{nr}|<\varepsilon} x_{nr} \left(e^{-\frac{x_{nr}^2}{2} + x_{nr} z_{ns}} - 1 \right) p_r \right\} (1+o(1)) =$$

$$= \frac{z_{ns}}{n} + \frac{o_\varepsilon(1)}{n} + \frac{O(\varepsilon)}{n}.$$

Аналогично получим

$$\alpha_2(\varepsilon, n, z_{ns}) = \frac{1}{n} + \frac{o_\varepsilon(1)}{n} + \frac{O(\varepsilon)}{n}.$$

Перейдем к доказательству теоремы 1. Для определенности опять предположим, что ξ_n — решетчатые величины. Прежде всего заметим, что если фиксировать величину $\eta_n = \eta_{nn}$, положив

$$\eta_{nn} = z_n, \tag{15}$$

то последовательность $\eta_{n0}, \eta_{n1}, \ldots, \eta_{nn}$ образует цепь Маркова. Таким образом, задача определения предельной условной вероятности события (5) при гипотезе (15) может быть решена методами, разработанными для цепей Маркова [10, 11].

Предположим, что задано $\eta_{nk} = s_n$, $\eta_{nn} = z_n$. Пользуясь леммой 1, получим для „урезанных“ условных моментов величины $\eta_{nk+1} - \eta_{nk} = \frac{\xi_{k+1}}{\sqrt{n}}$ следующие оценки:

$$1 - \alpha_0(\varepsilon, n, s_n, z_n) = \frac{o(1)}{n}, \quad \alpha_1(\varepsilon, n, s_n, z_n) = \left(\frac{z_n - s_n}{1 - t_k} + o_\varepsilon(1) + O(\varepsilon) \right) \frac{1}{n},$$

$$\alpha_2(\varepsilon, n, s_n, z_n) = (1 + o_\varepsilon(1) + O(\varepsilon)) \frac{1}{n},$$

где

$$t_k = \frac{k}{n} \leqslant T < 1.$$

Пользуясь предельными теоремами для последовательностей случайных величин, связанных в цепь Маркова (см., например, [11]) и принимая во внимание только-что полученные соотношения, непосредственно получим доказываемое.

Из доказанной теоремы легко получить предельную условную совместную функцию распределения пары функционалов $\{m_n(t), M_n(t)\}$ при гипотезе $\eta_n = z_n (n \to \infty, z_n \to z)$. Она равна [11, 12]

$$\Phi(a, b; t, z) = \frac{1}{\sqrt{2\pi}} \sum_{-\infty}^{\infty} \left\{ \exp\{-2r(b+a)[r(b+a)-z]\} \int_{a_{2r}}^{b_{2r}} e^{-u^2} du - \right.$$

$$\left. - \exp\{-2(rb + \overline{r-1}a)(rb + \overline{r-1}a - z)\} \int_{a_{2r-1}}^{b_{2r-1}} e^{-u^2} du \right\}, \tag{16}$$

где $a>0$, $b>0$,

$$
\begin{aligned}
a_{2r} &= \frac{-a-tz-2r(1-t)(b+a)}{\sqrt{t(1-t)}};\\
b_{2r} &= \frac{b-tz-2r(1-t)(b+a)}{\sqrt{t(1-t)}};\\
a_{2r-1} &= \frac{-a-tz-2(1-t)(rb+\overline{r-1}a)}{\sqrt{t(1-t)}};\\
b_{2r-1} &= \frac{b-tz-2(1-t)(rb+\overline{r-1}a)}{\sqrt{t(1-t)}}.
\end{aligned}
\tag{17}
$$

Предельное условное распределение величин

$$
M'(n, t) = \max_{0<k<nt} \frac{1}{\sqrt{n}}\left(\sum_1^k \xi_r - \frac{k}{n}\sum_1^n \xi_r\right),
$$

$$
m'(n, t) = -\min_{0<k<nt} \frac{1}{\sqrt{n}}\left(\sum_1^k \xi_r - \frac{k}{n}\sum_1^n \xi_r\right),
$$

при гипотезе $\frac{1}{\sqrt{n}}\sum\limits_1^n \xi_r = z_n\,(z_n \to z)$ равно $u(0, 0)$, где $u(t, s)$ есть решение уравнения (6), удовлетворяющее граничным условиям

$$
u(\tau, -a+\tau z) = u(\tau, b+\tau z) = 0, \quad 0 \leqslant \tau \leqslant t < 1,
$$
$$
u(t, s) = 1, \quad -a+tz < s < b+tz.
$$

Подстановка $y = s + tz$ преобразует уравнение (6) к тому же виду, где только нужно положить $z = 0$, а граничные условия будут

$$
u(\tau, -a) = u(\tau, b) = 0 \qquad 0 \leqslant \tau \leqslant t < 1,
$$
$$
u(t, s) = 0 \qquad -a < s < b.
$$

Таким образом, искомое условное распределение оказывается независящим от z и равно $\Phi(a, b, t, 0)$.

До сих пор мы считали $t < 1$. Предельные распределения для случая $t = 1$ можно получить используя теорему работы [11]. В силу этой теоремы

$$
\Phi(a, b;\, 1, z) = \lim_{t\to 1} \Phi(a, b;\, t, z).
$$

Теорема 2. Если последовательность случайных величин $\{\xi_n\}$ удовлетворяет условиям теоремы 1, z_n — в случае а) одно из возможных значений суммы

$$
\eta_n = \frac{1}{\sqrt{n}}\sum_1^n \xi_k, \quad n \to \infty \text{ и } z_n \to z,
$$

то 1) условная функция распределения пары величин

$$
\max_{0<k<n} \frac{1}{\sqrt{n}}\sum_1^k \xi_r, \quad -\min_{0<k<n} \frac{1}{\sqrt{n}}\sum_1^k \xi_r,
$$

при гипотезе $\eta_n = z_n$ стремится к

$$
\sum_{-\infty}^{\infty} \exp\{-2r(a+b)[r(a+b)-z]\} - \exp\{-2(rb+\overline{r-1}a)(rb+\overline{r-1}a-z\};
$$

2) предельная условная функция распределения величин

$$\max_{0<k<n} \frac{1}{\sqrt{n}}\left(\sum_1^k \xi_r - \frac{k}{n}\sum_1^n \xi_r\right), \quad -\min_{0<k<n} \frac{1}{\sqrt{n}}\left(\sum_1^k \xi_r - \frac{k}{n}\sum_1^n \xi_r\right),$$

при гипотезе $\eta_n = z_n$ не зависит от z и равна

$$1+\sum_1^{\infty} 2\exp\{-2r^2(a+b)^2\} - \exp\{-2(rb+\overline{r-1}a)^2\} - \exp\{-2(\overline{r-1}b+ra)^2\}.$$

Заметим еще, что в силу результатов уже цитированной работы [11] в теореме 1 можно рассматривать и неограниченные регулярные области, в частности можно обобщить теорему 1 и на области вида

$$\left\{-a_r < \eta_{nk_r} < b_r; \quad r=1, 2, \ldots, m\right\}, \tag{18}$$

где $\frac{k_r}{n} \to t_r$, $0 < t_r < 1$. Впрочем, последний результат может быть получен и непосредственно с помощью локальных предельных теорем для сумм независимых слагаемых. Предельное совместное условное распределение величин η_{n,k_r}, $\frac{k_r}{n} \to t_r$, $0<t_r<1$, при гипотезе $\eta_{nn} = z_n$, $z_n \to z$, совпадает с совместным распределением величин

$$\{\eta(t_1) + t_1 z, \ \eta(t_2) + t_2 z, \ \ldots, \ \eta(t_m) + t_m z\},$$

где $\eta(t)$ — случайная функция непрерывного марковского процесса, управляемого дифференциальным уравнением

$$\frac{\partial u}{\partial t} - \frac{x}{1-t}\frac{\partial u}{\partial x} + \frac{1}{2}\frac{\partial^2 u}{\partial x^2} = 0. \tag{19}$$

Случайный процесс $\{\eta(t)\}$ был подробно рассмотрен Дубом [13] в связи с эвристическим обоснованием теорем А. Н. Колмогорова [7] и Н. В. Смирнова [6]. Этот процесс полностью характеризуется следующими свойствами: а) m-мерная функция распределения величин $\eta(t_1), \eta(t_2), \ldots, \eta(t_m)$ нормальна, б) $M\eta(t_2)\eta(t_1) = t_1(1-t_2)$, $t_1 < t_2 < 1$. От сходимости условной вероятности на цилиндрических множествах легко можно перейти к сходимости по вероятности весьма общих функционалов от случайных функций. Мы рассмотрим ниже один пример такого рода, послуживший предметом специального исследования [4].

Пусть $f(x)$ непрерывная функция, определенная на $(-\infty, \infty)$, удовлетворяющая условию Липшитца

$$|f(x_1) - f(x_2)| \leqslant C\,|x_1 - x_2|.$$

Определим

$$I = \int_0^1 f[\eta(t)]\,dt$$

как предел в смысле сходимости в среднем по полуупорядоченному множеству разбиений сегмента [0, 1] интегральных сумм

$$\sum_{i=1}^n f[\eta(t_i')]\,\Delta t_i \tag{20}$$

$$\Delta t_i = t_i - t_{i-1}, \quad t_0 = 0, \quad t_n = 1, \quad t_{i-1} \leqslant t_i' \leqslant t_i,$$

если только этот предел существует.

Впрочем, существование этого интеграла для функций, удовлетворяющих условию Липшитца, устанавливается без труда. Если $\{\Delta t_{i,k}\}$ — некоторое подразбиение разбиения $\{\Delta t_i\}$, то

$$\delta = M\left|\sum_{i=1}^{n} f[\eta(t_i')]\Delta t_i - \sum_{i=1}^{n}\sum_{k=1}^{k_i} f[\eta(t'_{i,k})]\Delta t_{i,k}\right|^2 \leqslant$$

$$\leqslant C\sum_{i=1}^{n}\sum_{k=1}^{k_i} M|\eta(t_i') - \eta(t'_{i,k})|^2 \Delta t_{i,k}.$$

С другой стороны,

$$M|\eta(t_2) - \eta(t_1)|^2 = (t_2 - t_1)(1 - \overline{t_2 - t_1}),\ t_2 > t_1.$$

Отсюда следует

$$\delta \leqslant C\varepsilon, \tag{20'}$$

где $\varepsilon = \max\limits_{1 \leqslant i < n} \Delta t_i$. Таким образом, интегральные суммы (20) при $\varepsilon \to 0$ сходятся в смысле сходимости в среднем. Положим

$$\zeta_{nk} = \eta_{nk} - \frac{k}{n}\eta_{nn} \tag{21}$$

и рассмотрим суммы

$$I_n^* = \frac{1}{n}\sum_{k=1}^{n} f\left[\eta\left(\frac{k}{n}\right)\right],$$

$$I_{n,m} = \frac{1}{n}\sum_{k=1}^{n} f[\zeta_{nm,km}],\quad I_{n,1} = I_n.$$

Лемма 2. Если $f(x)$ удовлетворяет условию Липшитца и выполнены предположения теоремы 1, то условное распределение суммы I_n при гипотезе $\eta_n = z_n$ слабо сходится к распределению функционала I.

Введем характеристические функции $\psi(\lambda)$, $\psi_n^*(\lambda)$, $\psi_{n,m}(\lambda)$ величин I, I_n^*, $I_{n,m}$* соответственно. Тогда $\psi_n^*(\lambda) \to \psi(\lambda)$ при $n \to \infty$ равномерно относительно λ в любом конечном интервале $|\lambda| < A$. Пусть n — фиксировано. Имеем

$$|\psi(\lambda) - \psi_{nm}(\lambda)| \leqslant |\psi(\lambda) - \psi_n^*(\lambda)| + |\psi_n^*(\lambda) - \psi_{n,m}(\lambda)| + |\psi_{n,m}(\lambda) - \psi_{nm,1}(\lambda)|.$$

Для любого фиксированного $\varepsilon > 0$ подберем n таким, чтобы $|\psi(\lambda) - \psi_n^*(\lambda)| < \frac{\varepsilon}{3}$, $|\lambda| < A$. Тогда при $m \to \infty$

$$\psi_n^*(\lambda) - \psi_{n,m}(\lambda) \to 0.$$

Рассмотрим теперь разность $\psi_{n,m}(\lambda) - \psi_{nm,1}(\lambda)$. Имеем

$$|\psi_{n,m}(\lambda) - \psi_{nm,1}(\lambda)| \leqslant |\lambda| M\{g/\eta_{nm} = z_{nm}\},$$

где

$$g = \min\left\{2, \left|\frac{1}{n}\sum_{k=1}^{n} f(\zeta_{km}) - \frac{1}{nm}\sum_{k=1}^{nm} f(\zeta_k)\right|\right\}.$$

* $\psi_{n,m}(\lambda)$ — условная характеристическая функция величины $I_{n,m}$, соответствующая гипотезе $\eta_{nm} = z_{nm}$.

Так как

$$\left|\frac{1}{n}\sum_{k=1}^{n} f(\zeta_{km}) - \frac{1}{nm}\sum_{k=1}^{nm} f(\zeta_k)\right| \leqslant \frac{C}{nm}\sum_{k=1}^{n}\sum_{r=1}^{m} |\zeta_{km} - \zeta_{m(k-1)+r}|$$

и

$$|\zeta_{km} - \zeta_{m(k-1)+r}| \leqslant \frac{|z_{nm}|}{n} + \frac{1}{\sqrt{nm}}\left|\sum_{s=(k-1)m+r}^{km} \xi_s\right|,$$

то в силу леммы 1 $\left(\text{применяя ее к последовательности } \xi_k' = \frac{1}{\sqrt{m}}\sum_{(k-1)m+1}^{km} \xi_r\right)$ мы получим

$$M\{g|\eta_{nm} = z_{nm}\} \leqslant \frac{C}{nm}\sum_{k=1}^{n}\sum_{r=1}^{m}\left\{\frac{|z_{nm}|}{n} + O\left(\frac{1}{\sqrt{n}}\right)\right\} = O\left(\frac{1}{\sqrt{n}}\right).$$

Таким образом, равномерно относительно m разность $\psi_{n,m}(\lambda) - \psi_{nm}(\lambda)$ стремится к 0 при $n \to \infty$. Из этого следует

$$|\psi(\lambda) - \psi_{nm}(\lambda)| < \varepsilon$$

для всех достаточно больших $m > m_0(\varepsilon)$ при некотором фиксированном $n = n(\varepsilon)$. Но легко заметить, что

$$|\psi_N(\lambda) - \psi_{nm}(\lambda)| \to 0,$$

где $m = \left[\frac{N}{n}\right]$ — целая часть отношения $\frac{N}{n}$. Сопоставляя это с предыдущим неравенством, получим доказываемое.

Лемму, доказанную для функций $f(x)$, удовлетворяющих условию Липшитца, легко теперь обобщить на более общий класс функций.

Пусть $\{f_N^+(x)\}$, $\{f_N^-(x)\}$ — две монотонные последовательности функций, причем первая — монотонно убывающая, вторая — монотонно возрастающая, сходящиеся к некоторой функции $f_\infty(x)$. Положим $I^+(N) = I(f_N^+)$, $I^-(N) = I(f_N^-)$, и аналогично $I_n^+(N)$, $I_n^-(N)$. Тогда, если для функций f_N^+, f_N^- утверждение леммы имеет место, то

$$I_n^+(N) \geqslant I_n(\infty) \geqslant I_n^-(N), \quad I^-(N) \leqslant \underline{\lim}\, I_n(\infty) \leqslant \overline{\lim}\, I_n(\infty) \leqslant I^+(N).$$

Покажем теперь, что при $N \to \infty$ $I^+(N) - I^-(N)$ сходится в среднем к нулю. Из этого будет следовать, что последовательности интегралов $\{I^+(N)\}$, $\{I^-(N)\}$ при $N \to \infty$ сходятся в среднем к одному и тому же пределу, который можно определить как

$$I(f) = \int_0^1 f[\eta(t)]\,dt,$$

причем $\lim\limits_{n\to\infty} I_n(f)$ также существует и почти наверное совпадает с $I(f)$. Имеем

$$M|I^+(N) - I^-(N)|^2 \leqslant \int_0^1 M|f_N^+[\eta(t)] - f_N^-[\eta(t)]|^2\,dt.$$

Так как распределение $\eta(t)$ при фиксированном t абсолютно непрерывно относительно лебеговой меры на прямой, то из сходимости $f_N^+(x)-f_N^-(x)$ к нулю почти всюду на прямой следует также сходимость $f_N^+[\eta(t)]-f_N^-[\eta(t)]$ к нулю при $N\to\infty$ почти наверное. Применяя теорему Лебега о переходе к пределу под знаком интеграла, получим требуемое.

Мы доказали следующую лемму.

Лемма 3. Пусть $f(x)$ — произвольная ограниченная и измеримая по Лебегу функция; ζ_{nk} определяется соотношениями (21), и величины ξ_k удовлетворяют условиям теоремы 1.

Тогда условное распределение суммы

$$\frac{1}{n}\sum_{k=1}^{n} f[\zeta_{nk}]$$

при гипотезе

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}(\xi_k-M\xi_k)=z_n,$$

где z_n — допустимое (в случае решетчатых величин) значение суммы слева, и $z_n\to z$ при $n\to\infty$ слабо сходится к распределению функционала

$$I(f)=\int_0^1 f[\eta(t)]\,dt,$$

где $\eta(t)$ — случайная функция марковского процесса, управляемого уравнением (19).

Значение этой леммы состоит еще и в том, что она содержит в себе инвариантность предельного распределения по отношению к допустимым распределениям величин ξ_k. Так, например, если принять, что

$$f(x)=\begin{cases}1 & x>0\\ 0 & x\leqslant 0,\end{cases}$$

случайные величины $\xi_k=\pm 1$ с одинаковой вероятностью и $z=0$ (рассматриваемое нами распределение не зависит от z), то можно воспользоваться уже известными результатами об условном распределении числа положительных сумм $\sigma_k=\sum_1^k \xi_r$ при гипотезе $\sigma_n=0$ [1, 2]. Это приводит к следующей теореме.

Теорема 3. Пусть ν_n обозначает число положительных членов в последовательности $\left\{\sum_1^k \xi_r-\frac{k}{n}\sum_1^n \xi_r\right\}$ и ξ_k удовлетворяют условиям теоремы 1. Тогда при $n\to\infty$ условное распределение отношения

$$\frac{\nu_n}{n} \tag{22}$$

при гипотезе $\frac{1}{n}\sum_1^n \xi_r=z_n$, $z_n\to z$ слабо сходится к равномерному распределению.

Частный случай теоремы 3, соответствующий случаю $z=0$ и величинам ξ_k, принимающим значения, кратные 1, был недавно рассмотрен в уже упомянутой работе [4].

Так как предельное условное распределение не зависит от z, то безусловное распределение отношения (22) при $n\to\infty$ также сходится к равномерному распределению.

II

Предложенный А. Н. Колмогоровым в 1933 г. [7] критерий согласия между эмпирическими данными и заданной („теоретической“) функцией распределения $F(x)$

$$K_N = \max_{-\infty < x < \infty} |F_N(x) - F(x)|, \tag{1}$$

где $F_N(x)$ — эмпирическая функция распределения, построенная по N результатам наблюдений, является, как известно, универсальным в том смысле, что закон распределения величины K_N не зависит от непрерывной функции $F(x)$. Предельный закон распределения величины K_N $(N \to \infty)$ табулирован [14], имеются также таблицы распределений K_N для конечных N [15]. Во многих случаях, однако, относительно функции $F(x)$ известно только, что она принадлежит к некоторому типу распределений, зависящих от параметров θ, $F(x) = F(x, \theta)$, причем значения последних определяются эмпирически. В этом случае вместо величины (1) приходится вводить величину

$$\tilde{K}_N = \max_{-\infty < x} |F_N(x) - F(x, \bar{\theta})|, \tag{2}$$

где $\bar{\theta}$ — эмпирическое значение параметра θ, и положение резко меняется. Последнее связано с тем, что величина K_N при $N \to \infty$ неустойчива относительно параметра θ и в том случае, когда погрешность в определении параметра стремится к 0 вместе с N^{-1}. Одновременно с этим величина $\tilde{K}_N$ перестает быть универсальной, даже если метод оценки параметра стандартизован. Из результатов настоящей статьи следует, что в этом направлении простые закономерности отсутствуют.

Разобьем прямую $-\infty < x < \infty$ на n интервалов, назовем их интервалами группировки, и пусть $F_i(\theta)$ — вероятность случайной величине принять значение, лежащее в первых i интервалах, а $\bar{F}_i$ — эмпирическое распределение, т. е. отношение числа наблюдений, попавших в первые i интервалов группировки, к общему числу N произведенных наблюдений. Далее, пусть истинное значение параметра равно θ_0, а p_i — обозначает вероятность случайной величине принять значение, лежащее в i-том интервале группировки при значении параметра $\theta = \theta_0$. Положим

$$\xi_i = \sqrt{N}\{\bar{F}_i - F_i(\bar{\theta})\}, \tag{3}$$

$$\bar{\xi}_i = \sqrt{N}\left\{\bar{F}_i - F_i(\theta_0) - \left(\bar{\theta} - \theta_0;\ \frac{\partial F_i(\theta_0)}{\partial \theta_0}\right)\right\}, \tag{4}$$

$$\zeta_i = \bar{\xi}_i - \xi_i = \sqrt{N}\left\{F_i(\bar{\theta}) - F_i(\theta_0) - \left(\bar{\theta} - \theta_0;\ \frac{\partial F_i(\theta_0)}{\partial \theta_0}\right)\right\}, \tag{5}$$

где $(\theta_1; \theta_2)$ обозначает скалярное произведение двух m-мерных величин, а $\frac{\partial F_i(\theta_0)}{\partial \theta_0}$ — градиент функции $F_i(\theta)$ по переменной θ, взятый в точке θ_0.

Наша задача состоит в исследовании предельного распределения $\max\limits_{0<i<n} |\xi_i|$ при одновременном выполнении соотношений $N \to \infty$, $\max\limits_i p_i \to 0$.

Условимся называть две последовательности серий случайных величин

$$\xi_{\nu 1},\ \xi_{\nu 2}, \ldots,\ \xi_{\nu n_\nu} \tag{6}$$

$$\overline{\xi}_{\nu 1},\ \overline{\xi}_{\nu 2}, \ldots,\ \overline{\xi}_{\nu n_\nu}, \qquad \nu = 1, 2, \ldots \tag{7}$$

асимптотически эквивалентными, если при $\nu \to \infty$ предельные совместные распределения максимального и минимального членов в сериях (6) и (7) существуют одновременно, и если существуют, то совпадают.

Прежде всего установим следующую простую лемму.

Лемма 4. Если функция распределения $F(x, \theta)$ в некоторой окрестности точки θ_0 имеет ограниченные частные производные по θ, равномерно непрерывные относительно x, и если оценка $\overline{\theta}$ параметра θ такова, что величина

$$\sqrt{N}(\theta - \theta_0)$$

при $N \to \infty$ слабо сходится к некоторому пределу, то последовательности (3) и (4) асимптотически эквивалентны.

Доказательство этой леммы почти непосредственно вытекает из следующей.

Лемма 5. Если для последовательностей (6) и (7)

$$\max_{1 \leq i \leq n_\nu} |\xi_{\nu i} - \overline{\xi}_{\nu i}| \tag{8}$$

при $\nu \to \infty$ слабо сходится к 0, то эти последовательности асимптотически эквивалентны.

Доказательство. Положим

$$\mathrm{P}_\nu(a, b) = \mathrm{P}\{a < \xi_{\nu i} < b, \quad i = 1, 2, \ldots, n_\nu\},$$

$$\mathrm{P}(a, b) = \lim_{\nu \to \infty} \mathrm{P}_\nu(a, b),$$

и (a, b) — точка непрерывности функции распределения $\mathrm{P}(a, b)$.

Пусть

$$\overline{\mathrm{P}}_\nu(a, b) = \mathrm{P}\{a < \overline{\xi}_{\nu i} < b; \ \ i = 1, 2, \ldots, n_\nu\}.$$

Тогда

$$\overline{\mathrm{P}}_\nu(a, b) < \mathrm{P}_\nu(a - \varepsilon, b + \varepsilon) + \mathrm{P}\{\max_{1 \leq i \leq n_\nu} |\overline{\xi}_{\nu i} - \xi_{\nu i}| > \varepsilon\},$$

$$\mathrm{P}_\nu(a + \varepsilon', b - \varepsilon') < \overline{\mathrm{P}}_\nu(a, b) + \mathrm{P}\{\max_{1 \leq i \leq n_\nu} |\overline{\xi}_{\nu i} - \xi_{\nu i}| > \varepsilon'\},$$

где ε, ε' — произвольные положительные числа. Из этих неравенств и из условия (8) следует

$$\mathrm{P}(a + \varepsilon', b - \varepsilon') \leqslant \underline{\lim}\, \overline{\mathrm{P}}_\nu(a, b) \leqslant \overline{\lim}\, \overline{\mathrm{P}}_\nu(a, b) \leqslant \mathrm{P}(a - \varepsilon, b + \varepsilon).$$

Принимая во внимание непрерывность вероятности $\mathrm{P}(a, b)$ в рассматриваемой точке, получим требуемое.

Возвращаясь теперь к лемме 4, остается заметить, что с вероятностью, сколь-угодно близкой к 1,

$$|\zeta_i| = |\overline{\xi}_i - \xi_i| \leqslant N |\overline{\theta} - \theta_0|^2 \frac{C}{\sqrt{N}}.$$

В силу леммы 4 в дальнейшем можно вместо последовательности (3) ограничиться рассмотрением последовательности (4). Заметим еще, что в силу леммы 4 вместо величины $\sqrt{N}(\overline{\theta}-\theta_0)$ в (4) можно поставить, не изменяя предельного распределения $P(a, b)$, любую другую величину, отличающуюся от нее на бесконечно малую.

Если предположить, что оценка параметра θ производится по методу максимального правдоподобия, то (см., например, [16])

$$\sqrt{N'}\,(\overline{\theta}-\theta_0)=\frac{1}{l^2\sqrt{N'}}\sum_1^{N'}\frac{\partial\log f_i(\theta_0)}{\partial\theta}+\eta_{N'}=\frac{1}{\sqrt{N'}}\sum_1^{N'}\varphi_i+\eta_{N'},$$

где N' — число наблюдений, использованных для определения $\overline{\theta}$, $\eta_{N'}$ — величина, сходящаяся по вероятности к 0 при $N'\to\infty$, $f_i(\theta)$ — результат подстановки i-того наблюдения в плотность распределения $f(x, \theta)= =\dfrac{\partial F(x, \theta)}{\partial x}$ и

$$l^2=M\left(\frac{\partial\log f_i(\theta_0)}{\partial\theta}\right)^2.$$

В соответствии с этим в дальнейшем мы положим

$$\overline{\xi}_i=\sqrt{N}\left\{\overline{F}_i-F_i(\theta_0)-\frac{1}{\sqrt{N'}}\left(\Phi_{N'};\ \frac{\partial F_i(\theta_0)}{\partial\theta}\right)\right\}, \tag{9}$$

$$\Phi_{N'}=\frac{1}{\sqrt{N'}}\sum_{i=1}^{N'}\varphi_i, \tag{10}$$

где φ_i — независимые между собой, одинаково распределенные величины и

$$M\varphi_i=0,\quad M\varphi_i^2=\sigma^2<\infty.$$

Обычно для проверки заданной функции распределения $F(x)$ и для оценки неизвестного значения параметра θ используются одни и те же наблюдения. Однако не лишен значения и другой случай, когда для оценки параметра и проверки согласованности берутся два ряда независимых наблюдений. Остановимся прежде всего на второй возможности, исследование которой может быть произведено с помощью уже известных результатов, без затруднений.

Положим

$$\delta^2=\frac{N}{N'},\quad t=F(x, \theta_0),\quad t_i=F_i(\theta_0),\quad q(t)=\frac{\partial F(x(t), \theta_0)}{\partial\theta},$$

где $x(t)$ — функция обратная $t=F(x,\theta_0)$. Тогда

$$\overline{\xi}_i=\xi(t_i)=\sqrt{N}\{\overline{F}_i-t_i\}-\delta(\Phi_{N'};\ q(t_i)). \tag{11}$$

При этом мы предположим, что $q(t)$ непрерывно дифференцируемая функция t, и $q(0)=q(1)=0$.

Рассмотрим теперь дифференциальное уравнение

$$\frac{\partial u}{\partial\tau}+\frac{1}{2}\,\frac{\partial^2 u}{\partial x^2}=0;$$

пусть $u_T(\tau, x)$ — решение этого уравнения в области $0 \leqslant \tau < T$,

$$(1+\tau)\left[a+\delta_0\left(z;\, q\left(\frac{\tau}{1+\tau}\right)\right)\right] < x < (1+\tau)\left[b+\delta_0\left(z;\, q\left(\frac{\tau}{1+\tau}\right)\right)\right],$$

обращающееся в 0 на кривых

$$x = (1+\tau)\left[a+\delta_0\left(z;\, q\left(\frac{\tau}{1+\tau}\right)\right)\right],$$

$$x = (1+\tau)\left[b+\delta_0\left(z;\, q\left(\frac{\tau}{1+\tau}\right)\right)\right],$$

и удовлетворяющее условию $u_T(T, x) = 1$.

В силу принципа максимума для уравнения теплопроводности последовательность $u_T(\tau, x)$ при возрастании T монотонно убывает, и поэтому существует

$$\lim_{T\to\infty} u_T(\tau, x) = u_0(\tau, x).$$

Теорема 4. Если: 1) при $N \to \infty$ имеем $\frac{N}{N'} \to \delta_0^2$ и $\sqrt{N'}\,(\overline{\theta}-\theta_0)$ слабо сходится к некоторому пределу, 2) функция $F(x, \theta)$ непрерывна по x и обладает непрерывными и ограниченными частными производными по θ второго порядка, 3) оценка $\overline{\theta}$ параметра θ и эмпирическая функция распределения получены по двум последовательностям независимых между собой наблюдений, то условная вероятность совместного выполнения неравенств

$$a < \sqrt{N}\{\overline{F}_i - F_i(\overline{\theta})\} < b, \qquad i = 1, 2, \ldots, n, \tag{12}$$

при гипотезе

$$\sqrt{N'}\,(\overline{\theta}-\theta_0) = z_{N'},$$

где $z_{N'}$ сходится по вероятности к z, стремится к $u_0(0, 0) = u_0(0, 0/z)$.

Доказательство этой теоремы следует из результатов, полученных в [11]. При этом сходимость будет равномерной относительно z в любом конечном интервале $-A \leqslant z \leqslant A$. Из теоремы 4 легко получить предельную безусловную вероятность совмещения неравенств (12). Она равна

$$\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} u_0(0,\ 0/z)\, e^{-\frac{z^2}{2\sigma^2}}\, dz. \tag{13}$$

Перейдем теперь к более сложному случаю, когда эмпирическая функция распределения $\overline{F}_i$ и оценка параметра θ получены по одним и тем же результатам наблюдений. При этом $N' = N$.

В дальнейшем мы ограничимся случаем одного параметра. Основной результат, который будет ниже установлен, можно сформулировать следующим образом.

Теорема 5. Предположим, что:

a) оценка неизвестного параметра производится таким образом, что

$$\sqrt{N}\,(\overline{\theta}-\theta_0) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi(\xi_i) + \eta(N),$$

где $\{\xi_i\}$ — последовательность взаимно независимых измерений случайной величины ξ, случайная величина $\varphi(\xi)$ имеет конечный момент второго порядка, причем

$$M\varphi(\xi)=0, \quad M\varphi^2(\xi)=\sigma^2,$$

и $\eta(N)$ сходится по вероятности к 0;

b) к случайной величине $\psi(\xi)=\varphi(x(\xi))$, где $x(t)$ — функция, обратная функции $t=F(x, \theta_0)$, применима предельная теорема для плотностей;

c) функция $F(x, \theta)$ непрерывна по x и обладает непрерывными ограниченными производными первого и второго порядка.

Положим

$$\alpha(t)=\int_0^t \psi(\tau)\,d\tau, \quad q(t)=\frac{\partial}{\partial\theta}F(x(t), \theta)\Big|_{\theta=\theta_0},$$

$$A(t, s, \sigma/z)=\frac{[(1-t)(z-\sigma)-(s+zq(t))\alpha(t)][(1-t)\alpha'(t)+\alpha(t)]}{(1-t)\left[(1-t)\int_t^1 \alpha'^2(t)\,dt-\alpha^2(t)\right]}-$$

$$-\frac{s+zq(t)}{1-t}-zq'(t),$$

$$u_0(t, s, \sigma)=\lim_{\alpha\to 1} u_\alpha(t, s, \sigma),$$

где $u_\alpha(t, s, \sigma)$ — решение параболического уравнения

$$\frac{\partial u}{\partial t}+A(t, s, \sigma/z)\left(\frac{\partial}{\partial s}+\psi(t)\frac{\partial}{\partial\sigma}\right)u+\frac{1}{2}\left(\frac{\partial}{\partial s}+\psi(t)\frac{\partial}{\partial\sigma}\right)^2 u=0$$

в области $0\leqslant t<\alpha<1$, $-\infty<\sigma<+\infty$, $a_1<s<a_2$, $a_1<0<a_2$, удовлетворяющее условиям $u(\alpha, s, \sigma)=1$, $u(t, a_1, \sigma)=u(t, a_2, \sigma)=0$.

Тогда условная вероятность при гипотезе $\sqrt{N}(\overline{\theta}-\theta_0)=z_N$ совмещения неравенств

$$a_1<\sqrt{N}\{\overline{F}_i-F_i(\overline{\theta})\}<a_2, \qquad i=1, 2, \ldots, n,$$

стремится к $u_0(0, 0)$, когда $\max\limits_{1\leqslant i\leqslant n} p_i\to 0$ и z_N сходится по вероятности к z.

Пусть ξ — случайная величина с функцией распределения $F(x, \theta_0)$. Тогда $\tau_i=F(\xi_i, \theta_0)$ — случайная величина, равномерно распределенная на интервале (0,1). Введем случайную функцию

$$\chi(\tau_i, t)=\begin{cases}1, & \tau_i\leqslant t,\\ 0, & \tau_i>t.\end{cases}$$

Тогда в (11) мы можем положить

$$F_i=\frac{1}{N}\sum_{r=1}^{N}\chi(\tau_r, t_i),$$

$$\Phi_N=\frac{1}{\sqrt{N}}\sum_{r=1}^{N}\psi(\tau_r). \qquad (14)$$

Таким образом, $\delta = 1$

$$\bar{\xi}(t_i)=\sqrt{N}\left\{\frac{1}{N}\sum_{r=1}^{N}\chi(\tau_r, t_i)-t_i\right\}-\Phi_N q(t_i). \tag{15}$$

Имеем, далее,

$$\alpha(t)=M\chi(\tau_r, t)\psi(\tau_r)=\int_0^t \psi(\tau)\,d\tau, \quad \alpha(1)=0,$$

$$\beta(t)=D\chi(\tau_r, t)\,\psi(\tau_r)=\int_0^t \psi^2(\tau)\,d\tau-\alpha^2(t).$$

Введем еще последовательность величин

$$\bar{\eta}(t_i)=\sqrt{N}\left\{\frac{1}{N}\sum_{r=1}^{N}\chi(\tau_r, t_i)\psi(\tau_r)-\alpha(t_i)\right\}. \tag{16}$$

Очевидно,

$$\bar{\eta}(1)=\Phi_N.$$

Лемма 6. При гипотезе

$$\bar{\eta}(1)=\Phi_N=z$$

последовательность двумерных векторов

$$\{\bar{\xi}(t_i), \bar{\eta}(t_i); \quad i=1, 2,\ldots, n\}$$

связана в простую цепь Маркова.

Действительно, если $\bar{\eta}(1)=z$ задано, то, зная $\bar{\xi}(t_i)=s$, легко найти число случайных величин τ_r, имеющих значения, не превосходящие t_i. Обозначая его через $\nu_s(t_i)$, получим

$$\nu_s(t_i)=\sqrt{N}[s+zq(t_i)]+Nt_i.$$

Кроме того, если известно

$$\bar{\eta}(1)-\bar{\eta}(t_i),$$

то всякая дополнительная информация, которая может быть получена из знания величин $\bar{\eta}(t_j)$, $\bar{\xi}(t_j)$, $t_j<t_i$, сводится к информации о распределении результатов наблюдений, попавших в первые i интервалов группировки, и поэтому не оказывает дополнительного влияния на распределение величин $\bar{\xi}(t_j)$, $\bar{\eta}(t_j)$ при $j>i$.

Доказанная лемма открывает возможность изучения предельного условного распределения экстремальных членов последовательности $\{\bar{\xi}(t_i)\}$ с помощью метода дифференциальных уравнений [10, 11]. Для применения уже известных результатов нам понадобятся оценки для условных математических ожиданий приращений

$$\Delta\bar{\xi}(t_i)=\bar{\xi}(t_{i+1})-\bar{\xi}(t_i),$$

$$\Delta\bar{\eta}(t_i)=\bar{\eta}(t_{i+1})-\bar{\eta}(t_i)$$

при заданных

$$\bar{\xi}(t_i)=s, \quad \bar{\eta}(t_i)=\sigma, \quad \bar{\eta}(1)=z.$$

Рассмотрим теперь следующую вспомогательную задачу. Пусть $\{\tau_k\}$ — последовательность независимых равномерно распределенных на интервале (о, T) случайных величин. Положим

$$\eta_1(t)=\frac{1}{\sqrt{N}}\sum_{r=1}^{N}\left\{\chi(\tau_r,\ t)-\frac{t}{T}\right\},$$

$$\eta_2(t)=\frac{1}{\sqrt{N}}\sum_{r=1}^{N}\{\chi(\tau_r,\ t)\,\psi(\tau_r)-\alpha(t)\},$$

где

$$\alpha(t)=M\chi(\tau_r,\ t)\,\psi(\tau_r)=\frac{1}{T}\int_0^t\psi(\tau)\,d\tau,$$

$$\beta(t)=M\eta_2^2(t)=\frac{1}{T}\int_0^t\psi^2(\tau)\,d\tau-\alpha^2(t).$$

Требуется оценить условные моменты первых двух порядков величин $\eta_1(t)$, $\eta_2(t)$ для малых t при гипотезе $\eta_2(T)=z$. При этом мы будем предполагать, что $\psi(\tau_r)$ — абсолютно непрерывна (т. е. имеет плотность распределения) и удовлетворяет требованиям предельной теоремы для плотностей сумм независимых одинаково распределенных случайных величин. Тогда к $\eta_2(T)$ эта предельная теорема применима, и плотность распределения $p_2(T,z)$ величины $\eta_2(T)$ равномерно относительно z стремится к пределу

$$\frac{1}{\sqrt{2\pi\beta(T)}}\,e^{-\frac{z^2}{2\beta(T)}}.$$

Введем еще условную плотность $p_2(T,z\mid t,s,\sigma)$ величины $\eta_2(T)$ при гипотезе $\eta_1(t)=s$, $\eta_2(t)=\sigma$. Тогда, если $f(x,y)$ — произвольная непрерывная функция переменных x и y и если $M\{f[\eta_1(t),\eta_2(t)]/T,z\}$ — условное математическое ожидание величины $f[\eta_1(t),\eta_2(t)]$ при гипотезе $\eta_2(T)=z$, то из формулы Байесса следует

$$M\{f[\eta_1(t),\ \eta_2(t)]/T,\ z\}=Mf[\eta_1(t),\ \eta_2(t)]\,\gamma[\eta_1(t),\ \eta_2(t)], \tag{17}$$

где

$$\gamma(s,\ \sigma)=\frac{p_2(T,\ z|t,\ s,\ \sigma)}{p_2(T,\ z)}.$$

Займемся прежде всего величиною $\gamma(s,\sigma)$.

Величину $\eta_2(T)$ при гипотезе $\eta_1(t)=s$, $\eta_2(t)=\sigma$ можно отождествить с эквивалентной величиной

$$\eta_2(T)=\frac{1}{\sqrt{N}}\sum_{1}^{N-\nu_s}\psi(\tau_r')-\frac{T-t}{T}\sqrt{N}\,\alpha(t,\ T)+\sigma,$$

где

$$\alpha(t,\ T)=\frac{1}{T-t}\int_t^T\psi(\tau)\,d\tau,\qquad s=\frac{\nu_s}{\sqrt{N}}-\sqrt{N}\frac{t}{T}$$

и величины τ_r' равномерно распределены на интервале (t, T). Положим

$$\eta_2' = \frac{1}{\sqrt{N-\nu_s}} \sum_1^{N-\nu_s} [\psi(\tau_r') - M\psi(\tau_r')]; \tag{18}$$

тогда

$$\eta_2' = \sqrt{\frac{N}{N-\nu_s}} \left[\eta_2(T) - \sigma + \frac{s}{T-t} \int_t^T \psi(\tau)\, d\tau \right],$$

$$D\eta_2' = \beta(t, T) = \frac{1}{T-t} \int_t^T \psi^2(\tau)\, d\tau - \alpha^2(t, T).$$

В силу предельной теоремы для плотностей плотность распределения величины η_2' равномерно относительно t, $0 \leqslant t \leqslant T' < T$ стремится к

$$\frac{1}{\sqrt{2\pi\beta(t, T)}} e^{-\frac{z^2}{2\beta(t, T)}},$$

откуда

$$p_2(T, z | t, s, \sigma) \to \left(2\pi\beta(t, T)\frac{T-t}{T}\right)^{-\frac{1}{2}} \exp\left\{ -\frac{(z-\sigma+s\alpha(t, T))^2 T}{2\beta(t, T)(T-t)} \right\}$$

равномерно относительно $s, \sigma, -\infty < s, \sigma < +\infty$ и t в ранее указанном интервале. Отсюда уже следует

$$\gamma(s, \sigma) \to \left[\frac{\beta(0, T)T}{\beta(t, T)(T-t)}\right]^{\frac{1}{2}} \exp\left\{ -\frac{(z-\sigma+s\alpha(t, T))^2 T}{2\beta(t, T)(T-t)} + \frac{z^2}{2\beta(0, T)} \right\} = \gamma_0(t)$$

равномерно относительно t, s, σ и z, где $|z| < A$, A — любое > 0. В частности, величина $\gamma(s, \sigma)$ равномерно ограничена в указанных пределах.

При малых t

$$\gamma_0(t) = 1 + \frac{z(\sigma - s\alpha(0, T))}{\beta(0, T)} + O(t) + (\sigma^2 + s^2)\theta,$$

где θ — ограниченная функция.

Пусть $\chi_\varepsilon = \chi_\varepsilon(\eta_1, \eta_2)$ — функция, равная 1, если $|\eta_1| < \varepsilon$ и $|\eta_2| < \varepsilon$, и равная 0 в противном случае. Тогда для „урезанных“ моментов величин $\eta_1(t)$ и $\eta_2(t)$ без труда получим

$$M\{\chi_\varepsilon\eta_1(t) | T, z\} = \frac{z}{\beta(0, T)} M(\eta_1(t)\eta_2(t) - \alpha(0, T)\eta_1^2) + \delta_1(\varepsilon, t),$$

$$M\{\chi_\varepsilon\eta_2(t) | T, z\} = \frac{z}{\beta(0, T)} M(\eta_2^2(t) - \alpha(0, T)\eta_1(t)\eta_2(t)) + \delta_2(\varepsilon, t),$$

$$M\{\chi_\varepsilon\eta_1^2(t) / T, z\} = M\eta_1^2(t) + \delta_3(\varepsilon, t), \tag{19}$$

$$M\{\chi_\varepsilon\eta_1(t)\eta_2(t) | T, z\} = M\eta_1(t)\eta_2(t) + \delta_4(\varepsilon, t),$$

$$M\{\chi_\varepsilon\eta_2^2(t) | T, z\} = M\eta_2^2(t) + \delta_5(\varepsilon, t),$$

где величины $\delta_i(\varepsilon, t)$ таковы, что

$$\lim_{\varepsilon\to 0} \overline{\lim_{t\to 0}} \frac{\delta_i(\varepsilon, t)}{t} = 0.$$

Заметим еще, что из равномерной ограниченности величины $\gamma_0(t)$ следует

$$P\{\max[|\eta_1(t)|, |\eta_2(t)|] > \varepsilon/T, z\} \leqslant$$

$$\leqslant CP\{\max[|\eta_1(t)|, |\eta_2(t)|] > \varepsilon\} = o_\varepsilon(t), \tag{20}$$

где $o_\varepsilon(t)/t \to 0$, когда $t \to 0$ при любом фиксированном $\varepsilon > 0$. Для безусловных моментов величин $\eta_i(t)$ имеем следующие выражения:

$$M\eta_1(t) = M\eta_2(t) = 0, \tag{21}$$

$$M\eta_1^2(t) = \frac{t}{T}, \quad M\eta_1(t)\eta_2(t) = \frac{t}{T}\psi(0) + o(t), \quad M\eta_2^2(t) = \frac{t}{T}\psi^2(0) + o(t). \tag{22}$$

Возвратимся к величинам

$$\Delta\xi(t_i) = \sqrt{N}\left\{\frac{1}{N}\sum_1^{N_s}\chi(\tau_r', t_{i+1}) - \Delta t_i\right\} - \Phi_N \Delta q(t_i),$$

$$\Delta\eta(t_i) = \sqrt{N}\left\{\frac{1}{N}\sum_1^{N_s}\chi(\tau_r', t_{i+1}) - \Delta\alpha(t_i)\right\},$$

где τ_r' равномерно распределены на интервале $(t_i, 1)$ и

$$N_s = N - \nu_s = N(1 - t_i) - \sqrt{N}(s + zq(t_i)).$$

Имеем

$$\Phi' = \Phi_N - \eta(t_i) = \sqrt{N}\left\{\sum_1^{N_s}\frac{\psi(\tau_r')}{N} + \alpha(t_i)\right\} = z - \sigma$$

и

$$M\psi(\tau_r') = \frac{1}{1-t_i}\int_{t_i}^1 \psi(\tau)\,d\tau.$$

Положим

$$\tilde{\Phi} = \frac{1}{\sqrt{N_s}}\sum_1^{N_s}\psi(\tau_r') + \sqrt{N_s}\frac{\alpha(t_i)}{1-t_i};$$

тогда гипотеза $\Phi' = z - \sigma$ эквивалентна гипотезе $\tilde{\Phi} = \tilde{z}$, где

$$\tilde{z} = \sqrt{\frac{N}{N_s}}\left\{z - \sigma - \frac{s + zq(t_i)}{1-t_i}\alpha(t_i)\right\}. \tag{23}$$

Величина $\Delta\xi(t_i)$ может быть представлена в виде

$$\Delta\xi(t_i) = \sqrt{\frac{N_s}{N}}\left\{\frac{1}{\sqrt{N_s}}\sum_1^{N_s}\chi(\tau_r', t_{i+1}) - \sqrt{N_s}\frac{\Delta t_i}{1-t_i}\right\} -$$

$$-\frac{s + zq(t_i)}{1-t_i}\Delta t_i - z\Delta q(t_i). \tag{24}$$

Принимая во внимание формулы (19)—(22), получим для „урезанного" момента приращения $\Delta\xi(t_i)$ следующее выражение:

$$M\{\chi_\varepsilon \Delta\xi(t_i)/\tilde{\Phi}=\tilde{z}\} = \left(z-\sigma-\frac{s+zq(t)}{1-t}\alpha(t)\right)\frac{\psi(t)-\frac{1}{1-t}\int\limits_t^1\psi(\tau)d\tau}{(1-t)\beta(t)}\Delta t -$$

$$-\frac{s+zq(t)}{1-t}\Delta t - z\Delta q(t) + o_\varepsilon(\Delta t),$$

$$\beta(t) = \frac{1}{1-t}\int\limits_t^1 \psi^2(\tau)\,d\tau - \left(\frac{1}{1-t}\int\limits_t^1 \psi(\tau)\,d\tau\right)^2.$$

Аналогично получаются остальные „урезанные" моменты первого и второго порядка величин $\Delta\xi(t_i)$, $\Delta\eta(t_i)$. В силу уже цитированных результатов работы [11] и принимая во внимание замечания, сделанные непосредственно после формулировки теоремы, мы убеждаемся в справедливости теоремы 5.

ЛИТЕРАТУРА

1. Б. В. Гнеденко и В. С. Михалевич, ДАН СССР, 82, № 6, 1952.
2. K. L. Chung and W. Feller, Proc. Nat. Ac. Sc. 35 (1949).
3. M. Lipchitz, Proc. Am. Math. Soc., v. 3, № 4 (1952).
4. Б. В. Гнеденко и В. С. Михалевич, ДАН СССР, 85, № 1, 1952.
5. Б. В. Гнеденко и Е. Л. Рвачева, ДАН СССР, 82, № 4 (1952).
6. Н. В. Смирнов, Бюлл. Моск. гос. ун-та, в. 2 (1939).
7. А. Н. Колмогоров, Giornale Instit. Ital. Attuari, 4 (1933).
8. W. Feller, Ann. Math. Statistics, 22, № 3 (1951).
9. Б. В. Гнеденко и А. Н. Колмогоров, Предельные распределения для сумм независимых случайных величин.
10. А. Я. Хинчин, Асимптотические законы теории вероятностей, ОНТИ, 1936.
11. И. И. Гихман, Мат. сб. Киевского ун-та, № 8.
12. И. И. Гихман, ДАН СССР, 82, № 6, 1952.
13. J. L. Doob, Ann. Math. Statistics, 20, 3 (1949).
14. Б. В. Гнеденко, Курс теории вероятностей, ГИТТЛ, 1950.
15. Z. W. Birnbaum, Journal Am. St. Assoc., v. 47, № 259 (1952).
16. Г. Крамер, Математические методы статистики, ГИИЛ, 1948.

Получена 17 июня 1953 г.
Киев.