УДК: 519.234.2+519.237.5

П.Ф. Тарасенко

О СХОДИМОСТИ ИНДИКАТОРНЫХ ОЦЕНОК ДЛЯ ПАРАМЕТРОВ ЛИНЕЙНОЙ МОДЕЛИ

Рассматривается задача оценивания параметров линейной модели наблюдений, когда о распределении случайных погрешностей известно, что оно имеет определенные квантили заданных уровней. Введены оценки параметров по методу наибольшего достигнутого уровня значимости. Приводится доказательство равномерного закона больших чисел и состоятельности для одного класса индикаторных оценок. Предложенная оценка применима для привлечения априорной информации квантильного типа, а также для решения задач квантильной регрессии.

В некоторых приложениях статистических методов обработки данных встречаются ситуации, когда информация о случайных ошибках измерений формулируется в виде суждений об их принадлежности некоторым интервалам с заданной вероятностью. Это имеет место при обработке результатов физических опытов, когда экспериментатор не может делать утверждений о форме распределения погрешностей, но готов назвать точность измерений и формулирует это в виде интервала, которому принадлежит большая часть погрешностей. Похожая ситуация встречается при обработке вторичных данных статистического учета. При проведении статистических обследований принято контролировать точность определения показателей с помощью доверительных интервалов, при этом первичный материал не предоставляется в вышестоящие органы статистики, где решаются задачи анализа вторичных данных.

В самом общем виде модель этого типа априорной информации рассматривалась ранее в [2 – 4] применительно к задаче проверки гипотез о параметрах линейной модели.

В данной работе примем следующую модель погрешностей. Случайные величины $\varepsilon_1,\dots,\varepsilon_n$ независимы, одинаково распределены, а их общая функция распределения F имеет квантили $c_1,\dots,c_{K-1}$ уровней $q_1,\dots,q_{K-1}$ и непрерывна в точках $c_1\le\dots\le c_{K-1}$. Обозначим через $C_k=\langle c_{k-1},c_k\rangle$ (k=1,…,K) интервалы разбиения числовой оси, образованные квантилями. Здесь предполагается, что $c_0=-\infty$, $c_K=\infty$, а угловые скобки означают открытую или закрытую границу интервала. Каждому из интервалов приписана, таким образом, вероятность p_k, и если ф.р. F непрерывна на границах разбиения c_k, то $p_k=F(c_k)-F(c_{k-1})=q_k-q_{k-1}$.

На таком уровне априорной информации неизбежен отказ от использования величин невязок в качестве признаков и переход к использованию индикаторов попадания этих невязок в множества разбиения $\{C_k\}$. Очевидно, что такой переход осуществляется без потери различающей информации, поскольку о распределении погрешностей ничего не известно, кроме его квантилей.

В качестве основных примеров можно рассматривать две ситуации. В первом случае $\{c_k\}=\{-c, c\}$ и $\{p_k\}=\{p, 1-2p, p\}$. Во втором примере $\{c_k\}=\{-c,0,0,c\}$ и $\{p_k\}=\{p, 1-p, 0, 1-p, p\}$. Здесь постоянные $c>0$ и $p\in(0,{}^1/_2)$ будем считать известными. Случай неизвестного значения c, которое играет роль масштаба распределения, описывается в [2 – 4].

Случай известной медианы, приводящий к знаковому анализу, также может быть описан с помощью введенной модели априорных сведений о погрешностях, если положить $\{c_k\}=\{0,0\}$ и $\{p_k\}=\{{}^1/_2, 0, {}^1/_2\}$. В связи с этим техника доказательства результатов индикаторного статистического анализа во многом близка к доказательствам, полученным для знакового случая [1].

Со случаем $\{c_k\}=\{0,0\}$ и $\{p_k\}=\{p, 0, 1-p\}$ связан еще один класс задач, которые известны как квантильная регрессия (см., например, [5 – 7]). Традиционно методы квантильной регрессии основаны на М-оценке квантиля уровня p. В данной работе развивается другой подход к этой задаче, основанный на анализе индикаторов попадания невязок во множества разбиения $\{C_k\}$.

Следует заметить, что чем больше квантилей распределения погрешностей известно, тем лучшего качества можно ожидать от статистических выводов, если они используют знание этих квантилей. Таким образом, разработка индикаторных методов анализа данных может рассматриваться как стремление улучшить свойства статистических процедур за счет привлечения доступной априорной информации. Однако эта информация не выводит задачу за рамки непараметрического уровня априорной неопределенности.

В данной работе рассматривается один класс индикаторных оценок параметров линейной модели статистических наблюдений. Доказана состоятельность этих оценок.

ПОСТАНОВКА ЗАДАЧИ

Рассмотрим задачу оценивания параметров θ линейной модели

$$\boldsymbol{Y}=\boldsymbol{X}^T\theta+\varepsilon\,, \qquad (1)$$

которая описывает статистическую зависимость наблюдений $\boldsymbol{Y}=(Y_1,\dots,Y_n)^T$ от неизвестных параметров $\theta=(\theta_1,\dots,\theta_T)^T$, случайных погрешностей измерений $\varepsilon=(\varepsilon_1,\dots,\varepsilon_n)^T$ и матрицы плана $\boldsymbol{X}$, образованной столбцами $\boldsymbol{X}_1,\dots,\boldsymbol{X}_n$.

Оценка параметров θ может быть определена, если имеется метод проверки простой гипотезы H_0: $\theta=\theta_0$ против сложной альтернативы H_1: $\theta\ne\theta_0$. Заметим, что первичными признаками являются наблюдения $\boldsymbol{Y}$ (или соответствующие невязки $\boldsymbol{Y}-\boldsymbol{X}^T\theta_0$), однако статистическая проверка гипотез возможна только на основе использования априорного знания распределения $\{p_k\}$, которое задано на разбиении $\{C_k\}$. Это означает, что в качестве признаков достаточно использовать индикаторы попадания невязок модели (1) в множества разбиения $\{C_k\}$. Отсюда и происходит название индикаторных статистических процедур.

Оценкой параметров модели (1) может служить вектор θ_0, доставляющий наибольший достигнутый уровень значимости при проверке гипотезы H_0 против альтернативы H_1. В [2 – 4] предложено несколько вариантов тестов для проверки таких гипотез. Эти тесты уровня значимости γ представляются в виде $H(\boldsymbol{h}(\theta_0))>t_\gamma$, где $\boldsymbol{h}(\theta_0)=(h_1(\theta_0),\dots, h_n(\theta_0))^T$ – вектор индикаторных статистик $h_i(\theta_0)=s(Y_i-\boldsymbol{X}_i^T\theta_0)$, а индикаторная функция $s(u)$ принимает значение k, если $u\in C_k$.

Поскольку пороговое значение t_γ не зависит от θ_0, то оценка по методу максимума достигнутого уровня значимости запишется в виде

$$\Theta_n=\mathrm{U}_{s\in S_n}\Theta(\boldsymbol{s}),\ \text{где}\ S_n=\operatorname{Arg\,min}_{\mathbf{s}\in S(Y)}H(\boldsymbol{s}), \qquad (2)$$

$S(\boldsymbol{Y})$ – множество возможных наборов индикаторных признаков, $\Theta(\boldsymbol{s})$ – выпуклые многогранники в пространстве параметров, имеющие вид $\Theta(\boldsymbol{s})=\{\theta\text{: } \boldsymbol{h}(\theta)=\boldsymbol{s}\}$. При решении дискретной задачи минимизации (2) путем полного перебора, критерий $H(\boldsymbol{s})$ необходимо вычислять в K^n точках, поэтому для больших объемов выборки целесообразно использовать методы сокращения перебора, которые могут быть основаны на гипотезе «выпуклости» $H(\boldsymbol{h}(\theta))$ как функции от параметров θ.

В данной работе рассматриваются некоторые асимптотические свойства оценки (2), когда в качестве критериальной статистики используется функция

$$H(\boldsymbol{s})=\frac{1}{n}\sum_{j=1}^{T}\left(\sum_{i=1}^{n} X_{ij} B_1(s_i)\right)^2 . \quad (3)$$

Если ф.р. F непрерывно дифференцируема в окрестностях точек $c_1,\ldots,c_{K-1}$, то оптимальный выбор величин $B_1(k)$ предполагает (см. [4]) их задание в виде

$$B_1(k)=[f(c_k)-f(c_{k-1})]A/p_k,$$

где $f(c_k)=F'(c_k)$, $f(c_0)=f(c_K)=0$, A – произвольная постоянная. При необходимости вместо неизвестных величин $f(c_k)$ здесь могут быть использованы другие значения, играющие роль априорной догадки. Похожая ситуация с заданием весовых коэффициентов (меток) имеет место в ранговом анализе, где эти коэффициенты называются метками рангов. Поэтому веса $B_1(k)$ будем называть метками множеств разбиения $\{C_k\}$.

В частности, для двух основных примеров, описанных во введении, метки могут быть взяты в виде $\{B_1(k)\}=\{-1, 0, 1\}$ для первого примера, а для второго – в виде $\{B_1(k)\}=\{-2p/(1-2p),-\alpha, 0, \alpha, 2p/(1-2p)\}$. В последнем случае α – априорная догадка о значении величины $(f(0)-f(c))/f(c)$. Для случая квантильной регрессии получаем метки $\{B_1(k)\}=\{-1/p, 0, 1/(1-p)\}$.

В дальнейшем будем в разных сочетаниях использовать и ссылаться на следующие условия.

(а) Ф.р. F непрерывна и удовлетворяет условию Липшица, т.е $\exists\, L > 0$: $|F(u_1) - F(u_2)| < L\,|u_1 - u_2|$ $\forall u_1, u_2 \in R^1$.

(б) Матрица плана ограничена равномерно по n, т.е. $\exists H>0$: $\sup_{i,j,n}|X_{ij}|<H$.

(в) Метки $B_1(k)$ имеют нулевое среднее и не убывают, т.е.

$$\sum_{k=1}^{K} p_k B_1(k) = 0\,;\ B_1(k)\le B_1(k+1)\ \forall k=1,\ldots,K-1.$$

(г) Хотя бы для одного $k_0\in 1,..,K-1$ выполняется $B_1(k_0+1)>B_1(k_0)$ и для этого k_0 существуют постоянные $L_0>0$, $\delta>0$ такие, что $|F(u_1) - F(u_2)|\ge L_0|u_1-u_2|$ в δ-окрестности k_0-й границы априорного разбиения (т.е. при любых $u_1,u_2\in\{u:|u-c_{k0}|<\delta\}$).

(д) $B_1(s(u))\ge 0$ при $u>0$ и $B_1(s(u))\le 0$ при $u<0$.

(е) Матрица плана сходится и асимптотически не вырождена, т.е. существует предел

$$\lim_{n\to\infty} \boldsymbol{V}_{X,n} = \boldsymbol{V}_X >0, \text{ где } \boldsymbol{V}_{X,n}=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i\boldsymbol{X}_i^T .$$

РАВНОМЕРНЫЙ ЗАКОН БОЛЬШИХ ЧИСЕЛ

Введем обозначение для отклонения от истинных параметров $\boldsymbol{t}=\theta_0-\theta$ и переобозначим индикаторные признаки $s_i(\boldsymbol{t})=s(\varepsilon_i-\boldsymbol{X}_i^T\boldsymbol{t})=s(Y_i-\boldsymbol{X}_i^T\theta_0)=h_i(\theta_0)$, так что

$$P\{s_i(\boldsymbol{t})=k\}=$$
$$=\begin{cases} F(c_1+\boldsymbol{X}_i^T\boldsymbol{t}), & k=1,\\ F(c_k+\boldsymbol{X}_i^T\boldsymbol{t})-F(c_{k-1}+\boldsymbol{X}_i^T\boldsymbol{t}), & k=2,\ldots,K-1,\\ 1-F(c_{K-1}+\boldsymbol{X}_i^T\boldsymbol{t}), & k=K.\end{cases} \quad (4)$$

Теорема 1. Пусть выполняются условия (а), (б) и (в). Тогда для любого $A>0$ при всех $j=1,\ldots,T$ имеет место сходимость по вероятности:

$$\sup_{\|\boldsymbol{t}\|\le A}\left|\frac{1}{n}\sum_{i=1}^{n}X_{ij}B_1(s_i(\boldsymbol{t}))-\mathrm{M}\left\{\frac{1}{n}\sum_{i=1}^{n}X_{ij}B_1(s_i(\boldsymbol{t}))\right\}\right|\xrightarrow[n\to\infty]{P}0\,. \quad (5)$$

Доказательство. Зафиксируем j, обозначим через $U_n(\boldsymbol{t})$ выражение под знаком модуля в (5) и прежде всего докажем поточечную сходимость $U_n(\boldsymbol{t})$ к нулю по вероятности при любом фиксированном $\boldsymbol{t}$. Введем в рассмотрение центрированную величину $y_i(\boldsymbol{t})=B_1(s_i(\boldsymbol{t}))-\mathrm{M}\{B_1(s_i(\boldsymbol{t}))\}$. Заметим, что $\mathrm{M}\{U_n(\boldsymbol{t})\}=0$ и $\mathrm{M}\{U_n(\boldsymbol{t})\}\le B^2H^2/n\to 0$, где $B=\max_k|B_1(k)|$. Пользуясь неравенством Чебышева, заключаем, что $U_n(\boldsymbol{t})$ сходится к нулю по вероятности.

Для доказательства (5), в пространстве значений вектора $\boldsymbol{t}=(t_1,\ldots,t_T)^T$ опишем вокруг компакта $\{\boldsymbol{t}:\|\boldsymbol{t}\|\le A\}$ куб со стороной $2A$ и разобьем этот куб на $(2r)^T$ конгруэнтных кубов с границами-гиперплоскостями вида $t_k=Aj/r$ ($k=1,..,T$, $j=-r,..,r$). При всяких A и r существует конечное множество таких кубов, которые покрывают компакт $\{\boldsymbol{t}:\|\boldsymbol{t}\|\le A\}$. Обозначим это множество через $C(A,r)$ и рассмотрим произвольный куб $c\in C(A,r)$.

При фиксированном i линейная функция $\boldsymbol{X}_i^T\boldsymbol{t}$ достигает своего максимального и минимального значений внутри куба c на его диагонально противоположных вершинах. Величина $B_1(s_i(\boldsymbol{t}))=B_1(s(\varepsilon_i-\boldsymbol{X}_i^T\boldsymbol{t}))$ как функция от $\boldsymbol{X}_i^T\boldsymbol{t}$ является кусочно-постоянной и по условию (в) не возрастает. В результате получаем, что минимальные и максимальные значения функций $[-\boldsymbol{X}_i^T\boldsymbol{t}]$, $B_1(s_i(\boldsymbol{t}))$ и $\mathrm{M}\{B_1(s_i(\boldsymbol{t}))\}$ внутри куба c достигаются в одних и тех же точках – на диагонально противоположных вершинах этого куба, которые мы обозначим через $\boldsymbol{t}_{ic1}$ и $\boldsymbol{t}_{ic2}$ соответственно. При этом сами значения векторов $\boldsymbol{t}_{ic1}$ и $\boldsymbol{t}_{ic2}$ зависят только от вектора $\boldsymbol{X}_i$, поэтому они не являются случайными. Таким образом, для всех $\boldsymbol{t}\in c$ имеет место

$$A_1'-A_2'+A_3'-A_4'\le U_n(\boldsymbol{t})\le A_1-A_2+A_3-A_4,$$

где

$$A_1'=\frac{1}{n}\sum_{i=1}^{n}X_{ij}^{+}\left[B_1(s_i(\boldsymbol{t}_{ic1}))-\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic1}))\}\right],$$

$$A_1=\frac{1}{n}\sum_{i=1}^{n}X_{ij}^{+}\left[B_1(s_i(\boldsymbol{t}_{ic2}))-\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic2}))\}\right],$$

$$A_2'=\frac{1}{n}\sum_{i=1}^{n}X_{ij}^{-}\left[B_1(s_i(\boldsymbol{t}_{ic2}))-\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic2}))\}\right],$$

$$A_2=\frac{1}{n}\sum_{i=1}^{n}X_{ij}^{-}\left[B_1(s_i(\boldsymbol{t}_{ic1}))-\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic1}))\}\right],$$

$$A_3'=\frac{1}{n}\sum_{i=1}^{n}X_{ij}^{+}\left[\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic1}))\}-\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic2}))\}\right],$$

$$A_3=\frac{1}{n}\sum_{i=1}^{n}X_{ij}^{+}\left[\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic2}))\}-\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic1}))\}\right],$$

$$A_4'=\frac{1}{n}\sum_{i=1}^{n}X_{ij}^{-}\left[\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic2}))\}-\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic1}))\}\right],$$

$$A_4=\frac{1}{n}\sum_{i=1}^{n}X_{ij}^{-}\left[\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic2}))\}-\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic1}))\}\right],$$

$$X_{ij}^{+}=\max\{0,X_{ij}\},\ X_{ij}^{-}=-\min\{0,X_{ij}\}.$$

Сходимость к нулю по вероятности величин A_1, A_2, A_1', A_2' следует из неравенства Чебышева. Действительно, благодаря тому, что векторы $\boldsymbol{t}_{ic1}$ и $\boldsymbol{t}_{ic2}$ не являются случайными, все эти величины имеют нулевые средние, а их дисперсии ограничены сверху величиной $B^2H^2/n \to 0$.

Перейдем к рассмотрению величин A_3, A_4, A_3', A_4'. В силу условий теоремы, с учетом (4) справедлива оценка

$$\left|\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic2}))\} - \mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic1}))\}\right| \le$$
$$\le \sum_{k=1}^{K} \left|B_1(k)\right| 2L \left|\boldsymbol{X}_i^T \boldsymbol{t}_{ic2} - \boldsymbol{X}_i^T \boldsymbol{t}_{ic1}\right| \le 2KLBHAT/r,$$

где учтено неравенство $\|\boldsymbol{X}_i\| \le HT^{1/2}$ а также то, что точки $\boldsymbol{t}_{ic1}$ и $\boldsymbol{t}_{ic2}$ диагонально противоположны: $\|\boldsymbol{t}_{ic1} - \boldsymbol{t}_{ic2}\| = AT^{1/2}/r$. Отсюда, например, для A_3 получаем

$$|A_3| \le \frac{1}{n}\sum_{i=1}^{n} H \left|\mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic2}))\} - \mathrm{M}\{B_1(s_i(\boldsymbol{t}_{ic1}))\}\right| \le$$
$$\le 2KLBH^2AT/r.$$

В точности такая же оценка справедлива и для модулей величин A_4, A_3', A_4'.

Возвращаясь к (5), при всяком $\varepsilon>0$ выберем $r \ge KLBH^2AT/\varepsilon$ и рассмотрим событие Ω_n, состоящее в том, что для каждого $c \in C(A,r)$ величины A_1, A_2, A_1', A_2' по модулю не превосходят $\varepsilon/4$. В этих условиях событие $\max_{c\in C(A,r)} \sup_{\boldsymbol{t}\in c} |U_n(\boldsymbol{t})| \le \varepsilon$ является следствием события Ω_n и можно записать следующую цепочку неравенств:

$$P\left\{\sup_{\|\boldsymbol{t}\|\le A} |U_n(\boldsymbol{t})| \le \varepsilon\right\} \ge P\left\{\max_{c\in C(A,r)} \sup_{\boldsymbol{t}\in c} |U_n(\boldsymbol{t})| \le \varepsilon\right\} \ge P\{\Omega_n\}.$$

В силу доказанной ранее сходимости случайных величин A_1, A_2, A_1', A_2' к нулю по вероятности, при произвольных фиксированных r и ε, рассуждая от противного о совместном распределении этих величин, несложно убедиться, что $\lim_{n\to\infty} P\{\Omega_n\}=1$. Это и завершает доказательство теоремы 1.

СОСТОЯТЕЛЬНОСТЬ

Теорема 2. Пусть выполняются условия (а)–(е). Тогда оценка Θ_n, полученная по формулам (2) и (3), является состоятельной в том смысле, что выполняется равенство

$$\lim_{n\to\infty} P\{\sup_{\theta_n\in\Theta_n} \|\theta_n - \theta\| > \varepsilon\} = 0 \qquad (6)$$

Доказательство. Рассмотрим множества

$$T_{n,j} = \operatorname{Arg}\min_{\|\boldsymbol{t}\|\le A} |Z_j(\boldsymbol{t})|,$$

где $Z_j(\boldsymbol{t}) = \mathrm{M}\{\xi_j(\boldsymbol{t})\}$, $\xi_j(\boldsymbol{t}) = \frac{1}{n}\sum_{i=1}^{n} X_{ij} B_1(s_i(\boldsymbol{t}))$. (7)

Введем векторную функцию $\boldsymbol{Z}(\boldsymbol{t}) = (Z_1(\boldsymbol{t}),\ldots,Z_T(\boldsymbol{t}))$ и заметим, что по условию (в) имеем $\mathrm{M}\{B_1(s_i(\boldsymbol{0}))\}=0$ и, следовательно, $\boldsymbol{Z}(\boldsymbol{0})=\boldsymbol{0}$. Поэтому при любом n и j своего минимума, равного нулю, целевая функция $|Z_j(\boldsymbol{t})|$ из (7) достигает в точке $\boldsymbol{t}=\boldsymbol{0}$, т.е. $T_{n,j} = \{\boldsymbol{t}: Z_j(\boldsymbol{t})=0, \|\boldsymbol{t}\|\le A\}$ при всех n и $j=1,..,T$. Покажем, что при достаточно больших n множество $\{\boldsymbol{t}: \boldsymbol{Z}(\boldsymbol{t})=\boldsymbol{0}\}$ состоит из единственной точки. Предположим противное, пусть существует $\boldsymbol{t}\ne\boldsymbol{0}$, при котором $\boldsymbol{Z}(\boldsymbol{t})=\boldsymbol{0}$. В этом случае

$$\boldsymbol{t}^T \boldsymbol{Z}(\boldsymbol{t}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i^T \boldsymbol{t}\ \mathrm{M}\{B_1(s(\varepsilon_i - \boldsymbol{X}_i^T \boldsymbol{t}))\} = 0. \qquad (8)$$

Пользуясь (4), можно записать следующее выражение

$$\mathrm{M}\{B_1(s(\varepsilon_i - \boldsymbol{X}_i^T \boldsymbol{t}))\} =$$
$$= B_1(K) - \sum_{k=1}^{K-1} F(c_k + \boldsymbol{X}_i^T \boldsymbol{t})\left[B_1(k+1) - B_1(k)\right] =$$
$$= \sum_{k=1}^{K-1}\left[F(c_k) - F(c_k + \boldsymbol{X}_i^T \boldsymbol{t})\right]\left[B_1(k+1) - B_1(k)\right]. \qquad (9)$$

Если рассматривать это выражение как функцию от $\boldsymbol{X}_i^T\boldsymbol{t}$, то она не возрастает, а по условиям (в) и (г) для нее при если $\left|\boldsymbol{X}_i^T \boldsymbol{t} - c_{k_0}\right| < \delta$ справедлива оценка

$$\left|\mathrm{M}\{B_1(s(\varepsilon_i - \boldsymbol{X}_i^T \boldsymbol{t}))\}\right| \ge L_0 \left|\boldsymbol{X}_i^T \boldsymbol{t}\right| \left[B_1(k_0+1) - B_1(k_0)\right].$$

Это означает, что функция $\mathrm{M}\{B_1(s(\varepsilon_i - u))\}$ равна нулю только при $u=0$ и, следовательно, $u\mathrm{M}\{B_1(s(\varepsilon_i - u))\} \le 0$ при всех $u\in R^1$ и равенство здесь достигается только при $u=0$. Поэтому равенство в (8) возможно только если $\boldsymbol{X}_i^T\boldsymbol{t}=0$ для всех $i=1,..,n$. Однако в этом случае $\boldsymbol{t}^T \boldsymbol{V}_{X,n} \boldsymbol{t} = \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i \boldsymbol{t}\right)^2 = 0$ для данного $\boldsymbol{t}\ne 0$, что при больших n противоречит условию (е). Таким образом, $\{\boldsymbol{t}: \boldsymbol{Z}(\boldsymbol{t})=\boldsymbol{0}\} = \{\boldsymbol{0}\}$.

Рассмотрим далее вопрос об отделимости значений функции $\boldsymbol{Z}(\boldsymbol{t})$ от нуля за пределами окрестности точки $\boldsymbol{t}=\boldsymbol{0}$. Для этого, пользуясь (9), представим ее в виде

$$\boldsymbol{Z}(\boldsymbol{t}) = \sum_{k=1}^{K-1}\left[B_1(k+1) - B_1(k)\right] \times$$
$$\times \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \left[F(c_k) - F(c_k + \boldsymbol{X}_i^T \boldsymbol{t})\right].$$

Отсюда получаем

$$\boldsymbol{t}^T \boldsymbol{Z}(\boldsymbol{t}) = \sum_{k=1}^{K-1}\left[B_1(k+1) - B_1(k)\right] \times$$
$$\times \frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{X}_i^T \boldsymbol{t}\right)^2 \frac{\left[F(c_k) - F(c_k + \boldsymbol{X}_i^T \boldsymbol{t})\right]}{\boldsymbol{X}_i^T \boldsymbol{t}},$$

где слагаемые второй суммы, отвечающие случаю $\boldsymbol{X}_i^T\boldsymbol{t}=0$, считаются равными нулю. Применение к этому выражению условий (в) и (г) дает

$$-\boldsymbol{t}^T \boldsymbol{Z}(\boldsymbol{t}) \ge L_0\,[\,B_1(k_0+1) - B_1(k_0)\,]\,\boldsymbol{t}^T \boldsymbol{V}_{X,n} \boldsymbol{t},$$

если $\|\boldsymbol{t}\| < \delta/\max\|\boldsymbol{X}_i\|$. Из этого неравенства следует, что в указанной окрестности нуля

$$\|\boldsymbol{Z}(\boldsymbol{t})\| \ge \|\boldsymbol{t}\| \cdot L_0 \cdot [\,B_1(k_0+1) - B_1(k_0)\,] \cdot \min_{j=1,..,T} \lambda_j,$$

где λ_j – собственные значения матрицы $\boldsymbol{V}_{X,n}$. Таким образом, функция $\|\boldsymbol{Z}(\boldsymbol{t})\|$ непрерывна, и на любом компакте вида $K(A,B) = \{\boldsymbol{t}: B \le \|\boldsymbol{t}\| \le A\}$ выполняется $\|\boldsymbol{Z}(\boldsymbol{t})\| \ne 0$, если $\boldsymbol{V}_{X,n} > 0$ и $B>0$.

Отсюда следует, что функция $\|\boldsymbol{Z}(\boldsymbol{t})\|$ отделима от нуля на $K(A,B)$, т.е. $\forall A,B$ $(0<B<A)$ $\exists \varepsilon(A,B)>0$ такое, что $\forall \boldsymbol{t}\in K(A,B)$ имеет место $\|\boldsymbol{Z}(\boldsymbol{t})\| > \varepsilon(A,B)$, если $\boldsymbol{V}_{X,n}>0$. Действительно, если предположить обратное, то существует последовательность точек $\boldsymbol{t}_j \in K(A,B)$, для которой $\|\boldsymbol{Z}(\boldsymbol{t}_j)\| \to 0$ и из нее можно выбрать подпоследовательность, сходящуюся к некоторой точке $\boldsymbol{t}_0 \in K(A,B)$. В силу непрерывности функции $\|\boldsymbol{Z}(\boldsymbol{t})\|$ отсюда следует, что $\|\boldsymbol{Z}(\boldsymbol{t}_0)\| = 0$, и это противоречит установленному ранее факту.

Перейдем к рассмотрению индикаторной оценки (2) – (3). Обозначим $T_n = \{\boldsymbol{t}: \boldsymbol{t} = \theta_n - \theta,\ \theta_n \in \Theta_n\}$. Тогда

$T_n = \mathrm{Arg\,min}_{\boldsymbol{t}}\,\|\xi(\boldsymbol{t})\|$, где $\xi(\boldsymbol{t})=(\xi_1(\boldsymbol{t}),...,\xi_T(\boldsymbol{t}))^T$, $\xi_j(\boldsymbol{t})$ определены в (7). Условия теоремы (а)–(в) позволяют говорить о справедливости равномерного закона больших чисел (5), из которого следует, что

$$\sup_{\|\boldsymbol{t}\|\le A}\|\xi(\boldsymbol{t})-\boldsymbol{Z}(\boldsymbol{t})\| \underset{n\to\infty}{\overset{P}{\to}} 0.$$

Отсюда, в силу доказанной выше отделимости от нуля на компакте $K(A,B)$, для всех $B>0$ и $A>B$ выполняется

$$\lim_{n\to\infty} P\left\{\min_{B\le\|\boldsymbol{t}\|\le A}\left\|\boldsymbol{\xi}(\boldsymbol{t})\right\| > \tfrac{\varepsilon(A,B)}{2}\right\}=1. \qquad (10)$$

С другой стороны, для любого $\varepsilon > 0$ существует $B > 0$, при котором

$$\lim_{n\to\infty} P\left\{\sup_{\|\boldsymbol{t}\|<B}\left\|\boldsymbol{\xi}(\boldsymbol{t})\right\| < \varepsilon\right\}=1. \qquad (11)$$

Действительно, выполняется неравенство

$$\sup_{\|\boldsymbol{t}\|<B}\left\|\boldsymbol{\xi}(\boldsymbol{t})\right\| \le \sup_{\|\boldsymbol{t}\|<B}\left\|\boldsymbol{\xi}(\boldsymbol{t})-\boldsymbol{Z}(\boldsymbol{t})\right\| + \sup_{\|\boldsymbol{t}\|<B}\left\|\boldsymbol{Z}(\boldsymbol{t})\right\|,$$

причем первое слагаемое сходится по вероятности к нулю при любом $B>0$, а благодаря непрерывности функции $\|\boldsymbol{Z}(\boldsymbol{t})\|$ можно выбрать достаточно малое B так, чтобы второе слагаемое не превышало $\varepsilon/2$.

Теперь если в дополнение к (10) и (11) показать, что существует $A>0$ и $\varepsilon(A)>0$, при которых

$$\lim_{n\to\infty} P\left\{\min_{\|\boldsymbol{t}\|>A}\left\|\boldsymbol{\xi}(\boldsymbol{t})\right\| > \varepsilon(A)\right\}=1, \qquad (12)$$

то сходимость по вероятности (6) будет иметь место для случайного множества, заданного формулами (2) и (3), а доказательство теоремы будет завершено. Для этого заметим, что по условиям (в) и (д) $B_1(s_i(\boldsymbol{t}))\le 0$, если $\boldsymbol{X}_i^T\boldsymbol{t}<\varepsilon_i$; $B_1(s_i(\boldsymbol{t}))=0$, если $\boldsymbol{X}_i^T\boldsymbol{t}=\varepsilon_i$; $B_1(s_i(\boldsymbol{t}))\ge 0$, если $\boldsymbol{X}_i^T\boldsymbol{t}>\varepsilon_i$. Введем множества $I^+=\{i: |\boldsymbol{X}_i^T\boldsymbol{t}|>|\varepsilon_i|\}$, $I_+^-=\{i: |\boldsymbol{X}_i^T\boldsymbol{t}|\le|\varepsilon_i|,\ \varepsilon_i\cdot\boldsymbol{X}_i^T\boldsymbol{t}\ge 0\}$, $I_-^-=\{i: |\boldsymbol{X}_i^T\boldsymbol{t}|\le|\varepsilon_i|,\ \varepsilon_i\cdot\boldsymbol{X}_i^T\boldsymbol{t}<0\}$, $I^-=\{i: |\boldsymbol{X}_i^T\boldsymbol{t}|\le|\varepsilon_i|\}=I_+^-+I_-^-$. Заметим, что множества I^+, I_+^- и I_-^- образуют разбиение $\{1,...,n\}$ и имеет место $\{i: \boldsymbol{X}_i^T\boldsymbol{t}\cdot B_1(s_i(\boldsymbol{t}))>0\}\subset I^+ + I_-^-$, $\{i: \boldsymbol{X}_i^T\boldsymbol{t}\cdot B_1(s_i(\boldsymbol{t}))<0\}\subset I_+^-$. Воспользуемся неравенством $|a+b|\ge|a|-|b|$, чтобы получить оценку сверху для

$$\begin{aligned}
\left|\boldsymbol{t}^T\xi(\boldsymbol{t})\right| &= \left|\tfrac{1}{n}\textstyle\sum_{i=1}^n\left(\boldsymbol{X}_i^T\boldsymbol{t}\right)B_1(s_i(\boldsymbol{t}))\right| = \\
&= \left|\frac{1}{n}\sum_{i\in I^+ + I_-^-}\left|\left(\boldsymbol{X}_i^T\boldsymbol{t}\right)B_1(s_i(\boldsymbol{t}))\right| - \frac{1}{n}\sum_{i\in I_+^-}\left|\left(\boldsymbol{X}_i^T\boldsymbol{t}\right)B_1(s_i(\boldsymbol{t}))\right|\right| \ge \\
&\ge \frac{1}{n}\sum_{i\in I^+}\left|\left(\boldsymbol{X}_i^T\boldsymbol{t}\right)B_1(s_i(\boldsymbol{t}))\right| - \frac{1}{n}\sum_{i\in I_+^-}\left|\left(\boldsymbol{X}_i^T\boldsymbol{t}\right)B_1(s_i(\boldsymbol{t}))\right| = \\
&= \frac{1}{n}\sum_{i=1}^n\left|\left(\boldsymbol{X}_i^T\boldsymbol{t}\right)B_1(s_i(\boldsymbol{t}))\right| - \frac{2}{n}\sum_{i\in I_+^-}\left|\left(\boldsymbol{X}_i^T\boldsymbol{t}\right)B_1(s_i(\boldsymbol{t}))\right|.
\end{aligned} \qquad (13)$$

Оценим сверху последнее слагаемое в (13). Используем для этого соотношение, которое выполняется для любых неотрицательных a, b и c:

$$aI\{a<b\} = a\,I\{a<b\}[\,I\{b\le c\}+I\{b>c\}\,] \le$$
$$\le cI\{b\le c\}+a\,I\{b>c\},$$

где $I\{\cdot\}$ – индикатор выполнения условия. Тогда для любого $R>0$ имеет место оценка

$$\begin{aligned}
&\frac{2}{n}\sum_{i\in I_+^-}\left|\left(\boldsymbol{X}_i^T\boldsymbol{t}\right)B_1(s_i(\boldsymbol{t}))\right| \le \\
&\le \frac{2B_{1\max}}{n}\left[RI\{|\varepsilon_i|\le R\}+\left|\boldsymbol{X}_i^T\boldsymbol{t}\right|I\{|\varepsilon_i|>R\}\right] = \\
&= 2B_{1\max}\left[RP\{|\varepsilon_i|\le R\}+P\{|\varepsilon_i|>R\}\tfrac{1}{n}\textstyle\sum_{i=1}^n\left|\boldsymbol{X}_i^T\boldsymbol{t}\right|\right]+o_P(1),
\end{aligned}$$

где $B_{1\max}=\max_{k=1,K}|B_1(k)|$, а последнее соотношение получено по закону больших чисел. Продолжая цепочку (13), с помощью только что полученного соотношения можно утверждать, что для любых R и $\varepsilon>0$ неравенство

$$\begin{aligned}
\left|\boldsymbol{t}^T\xi(\boldsymbol{t})\right| &> \left[B_{1\min}-2B_{1\max}P\left\{\ |\varepsilon_i|>R\right\}\right]\times \\
&\times\tfrac{1}{N}\textstyle\sum_{i=1}^N\left|\boldsymbol{X}_i^T\boldsymbol{t}\right|-2RB_{1\max}P\left\{\ |\varepsilon_i|\le R\right\}-\varepsilon
\end{aligned} \qquad (14)$$

выполняется с вероятностью, которая сходится к единице при $n\to\infty$, где $B_{1\min}=\min_{k\in J}|B_1(k)|$, $J=\{k: B_1(k)\ne 0,\ k=1,..,K\}$. Чтобы полученное неравенство не было тривиальным, всегда можно выбрать достаточно большую постоянную R так, чтобы $B_{1\min}>2B_{1\max}Pr\{|\varepsilon_i|>R\}$.

Получим оценку снизу для величины $\frac{1}{n}\sum_{i=1}^n\left|\boldsymbol{X}_i^T\boldsymbol{t}\right|$, которая входит в (14). Для этого воспользуемся условием (е) и заметим, что

$$\tfrac{1}{n}\textstyle\sum_{i=1}^n\left|\boldsymbol{X}_i^T\boldsymbol{t}\right|^2 = \boldsymbol{t}^T\boldsymbol{V}_{X,n}\boldsymbol{t} \underset{n\to\infty}{\to} \boldsymbol{t}^T\boldsymbol{V}_X\boldsymbol{t} \ge \lambda\left\|\boldsymbol{t}\right\|^2,$$

где λ – наименьшее из собственных чисел положительно определенной матрицы $\boldsymbol{V}_X$. Поэтому, начиная с некоторого n, будет выполняться неравенство $\frac{1}{n}\sum_{i=1}^n\left|\boldsymbol{X}_i^T\boldsymbol{t}\right|^2 \ge \left\|\boldsymbol{t}\right\|^2\lambda/2$. С другой стороны, по условию (а) и неравенству Коши-Шварца, имеем $\frac{1}{n}\sum_{i=1}^n\left|\boldsymbol{X}_i^T\boldsymbol{t}\right|^2 \le H\left\|\boldsymbol{t}\right\|^2\frac{1}{n}\sum_{i=1}^n\left|\boldsymbol{X}_i^T\boldsymbol{t}\right|$. Объединяя полученные неравенства, приходим к оценке $\frac{1}{n}\sum_{i=1}^n\left|\boldsymbol{X}_i^T\boldsymbol{t}\right| \ge \left\|\boldsymbol{t}\right\|\lambda/(2H)$.

Используем полученное неравенство, чтобы продолжить цепочку (14). Итак, с вероятностью, стремящейся к единице с ростом n, при любых R и $\varepsilon>0$ выполняется неравенство

$$\left|\boldsymbol{t}^T\xi(\boldsymbol{t})\right| > \tilde{\lambda}\left\|\boldsymbol{t}\right\|-2RB_{1\max}P\left\{\ |\varepsilon_i|\le R\right\}-\varepsilon,$$

где $\quad\tilde{\lambda}=\left[B_{1\min}-2B_{1\max}P\left\{\ |\varepsilon_i|>R\right\}\right]\lambda/(2H)>0$.

Вернемся к доказательству (12). По неравенству Коши – Шварца и в силу полученных соотношений вероятность события

$$\|\boldsymbol{\xi}(\boldsymbol{t})\|^2 \ge \frac{\left[\boldsymbol{t}^T\xi(\boldsymbol{t})\right]^2}{\|\boldsymbol{t}\|^2} > \frac{\left[\tilde{\lambda}\|\boldsymbol{t}\|-2RB_{1\max}P\{\ |\varepsilon_i|\le R\}-\varepsilon\right]^2}{\|\boldsymbol{t}\|^2}$$

стремится к единице при любых R и $\varepsilon>0$. Правая часть этого неравенства стремится к $\tilde{\lambda}^2$ снизу при $\|\boldsymbol{t}\|\to\infty$. Поэтому существует достаточно большая постоянная A такая, что при всех $\|\boldsymbol{t}\|>A$ вероятность события $\|\boldsymbol{\xi}(\boldsymbol{t})\|^2>\tilde{\lambda}^2/2$ стремится к единице. Выбираемая величина A зависит от ε (возрастает с ростом ε), однако если рассматривать только достаточно малые $\varepsilon<\varepsilon_0$, то можно взять $A(\varepsilon)=A(\varepsilon_0)$. Этого достаточно, чтобы прийти к выводу о справедливости (12), что завершает доказательство теоремы 2.

Замечание 1. Условие (г) теоремы 2 можно усилить, заменив его следующей более простой формулировкой. Ф.р. F непрерывно дифференцируема в окрестностях точек $c_1,...,c_{K-1}$ и хотя бы для одного $k\in 1,..,K-1$ имеет место $F'(c_k)>0$ и $B_1(k+1)>B_1(k)$.

Замечание 2. Результаты теорем 1 и 2 можно распространить на более общий случай, когда вместо (3) при оценивании используется квадратичная форма вида $\boldsymbol{\xi}^T(\boldsymbol{t})\,\boldsymbol{W}_n\,\boldsymbol{\xi}(\boldsymbol{t})$. Для состоятельности такой оценки достаточно добавить условие $\lim_{n\to\infty}\boldsymbol{W}_n=\boldsymbol{W}>0$ и в соответствующих местах доказательства при получении верхних и нижних границ использовать неравенства, которые для сколь угодно малого w выполняются при достаточно больших n :

$$(w_{\min}-w)\,\|\boldsymbol{\xi}(\boldsymbol{t})\|^2\le w_{n,\min}\|\boldsymbol{\xi}(\boldsymbol{t})\|^2\le\boldsymbol{\xi}^T(\boldsymbol{t})\,\boldsymbol{W}_n\,\boldsymbol{\xi}(\boldsymbol{t})\le$$
$$\le w_{n,\max}\|\boldsymbol{\xi}(\boldsymbol{t})\|^2\le(w_{\max}+w)\,\|\boldsymbol{\xi}(\boldsymbol{t})\|^2.$$

Здесь $w_{n,\max}$ и $w_{n,\min}$ – величины максимального и минимального собственных значений матрицы $\boldsymbol{W}_n$, которые сходятся к собственным значениям $w_{\max}$ и $w_{\min}$ матрицы $\boldsymbol{W}$. В результате в формулах (10), (11) и (12) норма $\|\boldsymbol{\xi}(\boldsymbol{t})\|$ может быть заменена величиной $[\boldsymbol{\xi}^T(\boldsymbol{t})\,\boldsymbol{W}_n\,\boldsymbol{\xi}(\boldsymbol{t})]^{1/2}$. Соответствующие изменения произойдут с функцией $H(\boldsymbol{s})$, которая превратится в квадратичную форму. По поводу оптимального выбора матриц $\boldsymbol{W}_n$ см. [4].

Замечание 3. Для доказательства формулы (12) в теореме 2 использовано дополнительное условие (ж), которое не требуется для выполнения (10) и (11). Если имеют место только формулы (10) и (11), но не выполняется формула (12), то можно говорить о том, что при больших n в любой окрестности точки $\boldsymbol{t}=\boldsymbol{0}$ с вероятностью, сколь угодно близкой к единице, существует локальный минимум статистики $\boldsymbol{\xi}^T(\boldsymbol{t})\,\boldsymbol{W}_n\,\boldsymbol{\xi}(\boldsymbol{t})$. В связи с этим можно рассматривать множество локальных минимумов

$$T_n=\operatorname{Argu\,min}_{\boldsymbol{t}}\,\boldsymbol{\xi}^T(\boldsymbol{t})\,\boldsymbol{W}_n\,\boldsymbol{\xi}(\boldsymbol{t}),$$

где оператор Argu min выделяет множество точек, в которых достигаются локальные минимумы функции. Таким образом, если имеют место только условия (а) – (е), то для оценки

$$\Theta_n=\mathrm{U}_{\boldsymbol{s}\in S_n}\Theta(\boldsymbol{s}),\ \text{где}\ S_n=\operatorname{Argu\,min}_{\boldsymbol{s}\in S(\boldsymbol{Y})}H(\boldsymbol{s})$$

свойство состоятельности выполняется в форме

$$\lim_{n\to\infty}P\{\inf_{\theta_n\in\Theta_n}\|\theta_n-\theta\|>\varepsilon\}=0\,.$$

Замечание 4. Для примеров, рассмотренных во введении, из числа условий (а)-(е) выполняются все те, которые касаются меток $B_1(\cdot)$. Лишь для второго примера это требует дополнительного ограничения: $0<\alpha<2p/(1-2p)$. Таким образом, если последовательность матриц плана и функция распределения погрешностей также удовлетворяют всем ограничениям, то в рассмотренных примерах индикаторные оценки параметров являются состоятельными.

ЛИТЕРАТУРА

1. *Болдин М.В., Симонова Г.И., Тюрин Ю.Н.* Знаковый статистический анализ линейных моделей. М.: Наука, 1997.
2. *Tarassenko P.F.* Hypothesis testing for indicator analysis of linear models // KORUS-2002, 6th Russian-Korean International Symposium on Science and Technology, June 24-30, 2002, Novosibirsk: Proceedings. 2002. V.3, Mathematics. P.185.
3. *Tarassenko P.F.* On indicator-based hypothesis testing // 24th European Meeting of Statisticians (EMS2002), 14th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes – Prague, August 19-23, 2002.
4. *Тарасенко П.Ф.* Оптимальные тесты, основанные на индикаторах событий // Вестник ТГУ. 2002. Прил. №1(I), сент. Докл. IV Всеросс. конф. «Нов. информ. технологии в иссл. сложн. структур». Томск, 10 – 13 сентября 2002. С.185–190.
5. *Koenker R., Bassett G.* Regression quantiles // Econometrica. 1978. V.46. P.33–50.
6. *Koenker R., Portnoy S.* M Estimation of multivariate regressions // JASA. 1990. V.85, issue 412. P.1060–1068.
7. *Koenker R., Ng, P., Portnoy S.* Quantile smoothing splines // Biometrika. 1994. V.81. No. 4. P.673–680.

Статья представлена кафедрой теоретической кибернетики факультета прикладной математики и кибернетики Томского государственного университета, поступила в научную редакцию 30 апреля 2003 г.