

ПРОВЕРКА ЛИНЕЙНЫХ ГИПОТЕЗ НА ОСНОВЕ ИНДИКАТОРНЫХ ПРИЗНАКОВ

В рамках линейной модели независимых наблюдений предлагается и исследуется метод проверки линейных гипотез для ситуации, когда о распределениях случайных аддитивных погрешностей известно только, что они имеют общие квантили заданных уровней, известные с точностью до параметров. В частном случае такого рода априорная информация позволяет описывать неизвестный масштаб распределений погрешностей с помощью интерквантильного размаха. Приводятся тесты для проверки однородности выборок, основанные на индикаторах остатков.

Возможности построения статистических процедур во многом определяются имеющейся априорной информацией о стохастической составляющей обрабатываемых данных. В работе рассматривается класс квантильных априорных предположений о случайных погрешностях измерений. Они формулируются в виде суждений о квантилях распределения шумов, которые могут быть известны с точностью до параметров.

С одной стороны, такое описание априорной информации адекватно в ряде практических применений, например, при обработке результатов физических экспериментов и статистических обследований. При изучении зависимостей квантильная априорная информация приводит к описанию прогнозов в терминах квантилей условных распределений.

С другой стороны, знание квантилей с точностью до параметров слабее параметрического задания формы распределения. Такая информация оставляет задачу на непараметрическом уровне неопределенности. В то же время имеется возможность описывать неизвестный масштаб распределения с помощью интерквантильного размаха, известного с точностью до параметра.

Дадим формальное описание квантильной априорной информации. Пусть, для определенности, $Y_1, \dots, Y_n$ – скалярные количественные наблюдения, полученные в рамках линейной модели с аддитивными, случайными и независимыми погрешностями $\varepsilon_1, \dots, \varepsilon_n$. Модель наблюдений опишем позднее, а пока сосредоточимся на модели той априорной информации о распределениях погрешностей, с учетом которой мы собираемся строить статистические выводы. Пусть F_i – ф.р. с.в. ε_i и известно, что они могут различаться, но имеют общие квантили заданных уровней, известные с точностью до параметров. Иными словами, существуют векторы параметров μ_i , при которых

$$F_i(c_k + \mathbf{d}'_k \mu_i) = q_k, k = 1, \dots, K-1, i = 1, \dots, n, \quad (1)$$

где $0 < q_1 < \dots < q_{K-1} < 1$, $c_1 < \dots < c_{K-1}$ – известные величины, μ_i – M -мерный вектор неизвестных параметров, обычно имеющих смысл масштаба распределения, а векторы $\mathbf{d}_k$ описывают способ параметризации. При этом допустимыми являются значения параметров

$$\mu_i \in \{ \mu : c_1 + \mathbf{d}'_1 \mu < \dots < c_{K-1} + \mathbf{d}'_{K-1} \mu \}. \quad (2)$$

Считая ф.р. F_i непрерывной, $q_0=0$, $q_K=1$ и $c_0=-\infty$, $c_K=\infty$, обозначим через $p_k = q_k - q_{k-1}$ вероятности интервалов $C_k(\mu_i) = (c_{k-1} + \mathbf{d}'_{k-1} \mu, c_k + \mathbf{d}'_k \mu)$, а также введем векторы $\mathbf{p}=(p_1, \dots, p_K)'$, $\mathbf{c}=(c_1, \dots, c_{K-1})'$ и матрицу $\mathbf{D}=(\mathbf{d}_1, \dots, \mathbf{d}_{K-1})$. Тогда вектор границ интервалов примет вид $\mathbf{c} + \mathbf{D}' \mu_i$, а класс ф.р., удовлетворяющих при фиксированном μ_i априорным условиям (1), можно обозначить через $F_{K,M}(\mathbf{c} + \mathbf{D}' \mu_i, \mathbf{p})$. Поскольку вектор параметров μ_i неизвестен, то априорная информация состоит в том, что при известных $\mathbf{c}$, $\mathbf{D}$ и $\mathbf{p}$ выполняется

$$F_i \in \{ F : F_{K,M}(\mathbf{c} + \mathbf{D}' \mu_i, \mathbf{p}), \mu_i \in M_{K,M}(\mathbf{c}, \mathbf{D}) \},$$

где $M_{K,M}(\mathbf{c}, \mathbf{D})$ – множество допустимых параметров, определенное в (2).

В качестве основных примеров будем рассматривать и далее ссылаться на следующие ситуации.

Пример 1. Если требуется описать масштаб шумов с помощью интерквантильного размаха уровня p , то достаточно

положить $K=3$, $M=1$, $\mathbf{c}=(-c, c)'$, $\mathbf{D}=(-1, 1)'$ и $\mathbf{p}=(p, 1-2p, p)'$, где постоянные $c \geq 0$ и $p \in (0, 1/2)$ известны. Здесь $M_{K,M}(\mathbf{c}, \mathbf{D}) = (-c, \infty)$ и предполагается, что середина интерквантильного интервала находится в нуле.

Пример 2. Если дополнительно известна медиана распределения шумов, то полагаем $K=4$, $M=1$, $\mathbf{c}=(-c, 0, c)'$, $\mathbf{D}=(-1, 0, 1)'$, $\mathbf{p}=(p, 1/2-p, 1/2-p, p)'$ при заданных постоянных $c > 0$ и $p \in (0, 1/2)$.

Пример 3. Еще с одним классом задач [1–3], который известен как квантильная регрессия, связан случай, когда $K=2$, $M=0$, $\mathbf{c}=0$, $\mathbf{p}=(p, 1-p)'$. Здесь единственная общая для всех шумов квантиль известна, поэтому отсутствуют параметры априорного разбиения.

Пример 4. В частном случае, если у распределения шумов общая медиана, то $K=2$, $M=0$, $\mathbf{c}=0$, $\mathbf{p}=(1/2, 1/2)'$. Этот класс задач рассматривается, например, в знаковом анализе [4].

Задачи проверки гипотез и оценивания параметров линейной модели наблюдений в условиях сформулированной выше модели шумов рассматривались ранее в [5–7]. В данной работе строятся и исследуются процедуры проверки линейных гипотез в рамках линейной модели наблюдений.

ПОСТАНОВКА ЗАДАЧИ

Рассмотрим линейную модель наблюдений

$$Y_i = X'_i \theta_i + \varepsilon_i, \quad i = 1, \dots, n, \quad (3)$$

которая описывает статистическую зависимость наблюдений $\mathbf{Y} = (Y_1, \dots, Y_n)'$ от неизвестных параметров $\theta_i = (\theta_{1i}, \dots, \theta_{Ti})'$, независимых случайных погрешностей измерений $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ и матрицы плана $\mathbf{X}$, образованной столбцами $\mathbf{X}_1, \dots, \mathbf{X}_n$. При этом для векторных параметров $\theta_1, \dots, \theta_n$ заданы линейные ограничения-равенства. Например, в задачах с двумя внутренне однородными выборками объемов k и $n-k$ имеем

$$\theta_1 = \dots = \theta_k, \quad \theta_{k+1} = \dots = \theta_n.$$

Будем рассматривать случай, когда о распределении этих величин ε_i (ф.р. F_i) известно только, что

$$F_i \in F_{K,M}(\mathbf{c} + \mathbf{D}' \mu_i, \mathbf{p}), \quad (4)$$

причем векторы $\mu_i \in M_{K,M}(\mathbf{c}, \mathbf{D})$ неизвестны, но могут быть априори линейно связаны. Например, если две выборки объемов k и $n-k$ могут различаться по масштабу шумов, то $\mu_1 = \dots = \mu_k$, $\mu_{k+1} = \dots = \mu_n$.

В общем случае объединенный вектор параметров μ_i и θ_i априори принадлежит некоторому линейному подпространству L , $\dim L = m \leq n(T+M)$, а линейная гипотеза выдвигается относительно принадлежности объединенного вектора некоторому подпространству $L_0 \subset L$. Обозначим через L и L_0 соответственно матрицы базисных векторов-строк пространств L и L_0 , и пусть их размерности равны

$$\dim L = m \times n(T+M), \quad \dim L_0 = m_0 \times n(T+M).$$

Пусть линейное подпространство L_1 есть некоторое дополнение L_0 до L и L_1 – матрица его базисных векторов-строк, $\dim L_1 = (m-m_0) \times n(T+M)$. Тогда, если иметь

в виду очевидные обозначения для блоков введенных матриц, то пространство L можно параметризовать

$$\theta_i = L'_{01i}\omega + L'_{11i}\tau, \quad \mu_i = L'_{02i}\omega + L'_{12i}\tau, \quad i = 1, \dots, n,$$

а линейная гипотеза и ее альтернатива примут вид

$$H_0: \tau = 0, H_1: \tau \neq 0. \quad (5)$$

Введенная модель пригодна для описания линейных гипотез относительно параметров, однако чаще используется более удобный и общий способ выдвижения линейных гипотез об откликах $Y_1, \dots, Y_n$, а не о параметрах θ_i . Это приводит к модели

$$Y_i = L'_{1i}\omega + J'_{1i}\tau + \varepsilon_i, \quad i = 1, \dots, n, \quad (6)$$

$$F_i \in F_{K,M}(\mathbf{c} + (L_{2i}\mathbf{D})'\omega + (J_{2i}\mathbf{D})'\tau, \mathbf{p}),$$

причем $L'_{2i}\omega + J'_{2i}\tau \in M_{K,M}(\mathbf{c}, \mathbf{D})$, расширенный вектор наблюдений $\mathbf{Z} = (Y_1, \dots, Y_n, \mu'_1, \dots, \mu'_n)'$ априори принадлежит некоторому линейному пространству L , $\dim L = m \leq n(1+M)$, линейная гипотеза есть $\mathbf{Z} \in L_0 \subset L$, матрица $\mathbf{L} = (L_{11}, \dots, L_{1n}, L_{21}, \dots, L_{2n})$ составлена из базисных векторов-строк пространства L_0 , $\dim L_0 = m_0 < m$, $\mathbf{J} = (J_{11}, \dots, J_{1n}, J_{21}, \dots, J_{2n})$ – базисная матрица пространства L_1 , дополняющего L_0 до L . В конкретных постановках задач матрицы $\mathbf{A}$ и $\mathbf{L}$ базисных векторов-строк пространств L и L_0 определяются содержательным смыслом. При этом базис $\mathbf{J}$ в L_1 не обязательно ортогонален $\mathbf{L}$, способ его выбора, который будет определен позднее, диктуется стремлением обеспечить желательные свойства теста для проверки линейных гипотез вида (5) по наблюдениям $\mathbf{Y}$.

СИНТЕЗ ТЕСТА

Будем строить статистические выводы на основе фактов (индикаторов) принадлежности остатков модели (6) множествам априорного разбиения. Для этого перейдем от остатков к их индикаторам

$$s_i = s_i(\omega_0, \tau_0) = s(Y_i - L'_{1i}\omega_0 - J'_{1i}\tau_0, L'_{2i}\omega_0 - J'_{2i}\tau_0),$$

где индикаторная функция $s(u, \mu_0)$ принимает значение k , если $u \in C_k(\mu_0) = (c_{k-1}(\mu_0), c_k(\mu_0))$, где $c_k(\mu_0) = c_k + \mathbf{d}'_k \mu_0$ – гипотетическая квантиль уровня q_k для распределений шумов. В предположении, что ф.р. F_i непрерывны, запишем правдоподобие $P\{\mathbf{s} | \omega, \tau\}$ индикаторных признаков $\mathbf{s} = (s_1, \dots, s_n)'$:

$$P\{\mathbf{s}(\omega_0, \tau_0) | \omega, \tau\} = \prod_{i=1}^n P\{s_i(\omega_0, \tau_0) | \omega, \tau\},$$

где

$$P\{s_i(\omega_0, \tau_0) = k | \omega, \tau\} =$$

$$= F_i(c_k(\mu_i(\omega_0, \tau_0)) - L'_{1i}(\omega - \omega_0) - J'_{1i}(\tau - \tau_0) | \mu_i(\omega, \tau)) -$$

$$- F_i(c_{k-1}(\mu_i(\omega_0, \tau_0)) - L'_{1i}(\omega - \omega_0) - J'_{1i}(\tau - \tau_0) | \mu_i(\omega, \tau)),$$

$\mu_i(\omega_0, \tau_0) = L'_{2i}\omega_0 + J'_{2i}\tau_0$, обозначение $F_i(x | \mu_i)$ не означает, что F_i известна с точностью до параметров μ_i , а только подчеркивает, что она принадлежит классу $F_{K,M}(\mathbf{c} + \mathbf{D}'\mu_i, \mathbf{p})$, т.е. выполняется (4). Однако можно рассматривать производные $F_i(x | \mu_i)$ по μ_i на произвольно параметризованном семействе альтернативных распределений, на котором $F_i(x | \mu_i)$ становится непрерывной функцией переменных x и μ_i . Такое параметризованное семейство будем называть траекторией альтернативных распределений и обозначать его через $F_{K,M}(\mathbf{c}, \mathbf{D}, \mathbf{p})$. Пусть на такой траектории $F_i(x | \mu_i)$ дифференцируема по x , μ_i и имеет непрерывную плотность $f_i(x | \mu_i)$ в окрестностях точек $(x, \mu_i) = (c_k(\mu_i), \mu_i)$. Благодаря (4), для всех таких траекторий производные правдоподобия по параметрам ω и τ в гипотетической точке одинаковы:

$$\begin{aligned} \nabla P\{s_i(\omega_0, \tau_0) | \omega = \omega_0, \tau = \tau_0\} = \\ = -P\{s_i(\omega_0, \tau_0) | \omega_0, \tau_0\} \left(\begin{matrix} \mathbf{L} \\ \mathbf{J} \end{matrix} \right) \mathbf{R}(\mathbf{s}(\omega_0, \tau_0)), \end{aligned}$$

где

$$\mathbf{R}(\mathbf{s}) = (R_1(s_1), \dots, R_1(s_n), R'_2(s_1), \dots, R'_2(s_n))',$$

$$R_1(k) = [f_{i,k} - f_{i,k-1}] / p_k,$$

$$R'_2(k) = [\mathbf{d}'_k f_{i,k} - \mathbf{d}'_{k-1} f_{i,k-1}] / p_k,$$

$f_{i,k} = f_i(c_k(\mu_i) | \mu_i)$. Это позволяет строить статистические выводы на основе градиента отношения правдоподобия

$$\nabla_{(\omega, \tau)} \left[\frac{P\{s_i(\omega_0, \tau_0) | \omega, \tau\}}{P\{s_i(\omega_0, \tau_0) | \omega_0, \tau_0\}} \right]_{\substack{\omega = \omega_0 \\ \tau = \tau_0}} = - \left(\begin{matrix} \mathbf{L} \\ \mathbf{J} \end{matrix} \right) \mathbf{R}(\mathbf{s}(\omega_0, \tau_0)). \quad (7)$$

Для этого, прежде всего, наборы неизвестных величин $R_j = \{R_j(1), \dots, R_j(K)\}$ следует заменить наборами $B_j = \{B_j(1), \dots, B_j(K)\}$, которые называются метками. Часто B_j удается выбрать пропорциональными R_j , полностью сохранив смысл метода максимального отношения правдоподобия. В противном случае метки B_j трактуются как полученные в результате априорных догадок о величинах $f_{i,k}$. После такой замены должно выполняться свойство $\sum_{k=1}^K B_j(k) p_k = 0$.

Обозначим моменты меток через $d_1^2 = \sum_{k=1}^K B_1^2(k) p_k$,

$$\mathbf{D}_2 = \sum_{k=1}^K \mathbf{B}_2(k) \mathbf{B}_2'(k) p_k, \quad \mathbf{C} = \sum_{k=1}^K \mathbf{B}_1(k) \mathbf{B}_2(k) p_k.$$

Процедура проверки линейной гипотезы (5) будет состоять из двух шагов. На первом шаге, в предположении, что гипотеза выполняется, получим оценки параметров ω . На втором шаге, заменив истинное значение ω полученной оценкой, проверим гипотезу (5). Если на обоих этапах использовать принцип максимума отношения правдоподобия (7), то полученная процедура запишется в виде.

$$\omega_n = \underset{\omega_0 \in \Omega}{\operatorname{argmin}} \mathbf{B}'_n(\omega_0, \mathbf{0}) \mathbf{L}' \mathbf{W}_{1n} \mathbf{L} \mathbf{B}_n(\omega_0, \mathbf{0}); \quad (8)$$

$$\mathbf{B}'_n(\omega_n, \mathbf{0}) \mathbf{J}' \mathbf{W}_{2n} \mathbf{J} \mathbf{B}_n(\omega_n, \mathbf{0}) > t_\gamma, \quad (9)$$

где статистика $\mathbf{B}_n(\omega_0, \tau_0) = \mathbf{B}(\mathbf{s}(\omega_0, \tau_0))$ определена как $\mathbf{B}(\mathbf{s}) = n^{-1/2} (B_1(s_1), \dots, B_1(s_n), B'_2(s_1), \dots, B'_2(s_n))'$, $\mathbf{W}_{1n}$ и $\mathbf{W}_{2n}$ образуют последовательности положительно определенных и невырожденных (в том числе в пределе) матриц, Ω – множество допустимых значений параметров, т.е.

$$\Omega = \{\omega: L'_{2i}\omega \in M_{K,M}(\mathbf{c}, \mathbf{D}), i = 1, \dots, n\}.$$

В качестве t_γ можно взять квантиль уровня γ предельного распределения с.в. $\mathbf{B}'(\mathbf{s}) \mathbf{J}' \mathbf{W}_{2n} \mathbf{J} \mathbf{B}(\mathbf{s})$ или использовать квантиль точного распределения этой с.в., которую можно получить методом Монте-Карло, если иметь в виду, что $P\{s_i = k\} = p_k$. В нормированном варианте $\mathbf{W}_{2n} = (\mathbf{J} \mathbf{V}_B \mathbf{J}')^{-1}$, где

$$\mathbf{V}_B = \operatorname{var}\{\mathbf{B}_n(\mathbf{0}, \mathbf{0})\} = \begin{pmatrix} d_1^2 \mathbf{I}_n & \mathbf{e}_1 \mathbf{C}' & \mathbf{e}_n \mathbf{C}' \\ \mathbf{C} \mathbf{e}'_1 & \mathbf{D}_2 & \mathbf{0} \\ \mathbf{C} \mathbf{e}'_n & \mathbf{0} & \mathbf{D}_2 \end{pmatrix},$$

$\mathbf{I}_n = (\delta_{ij} : i, j = 1, \dots, n) = (\mathbf{e}_1, \dots, \mathbf{e}_n)$ – единичная матрица,

$\mathbf{e}_j = (\delta_{ij} : i = 1, \dots, n)$. В этом случае можно положить

$t_\gamma = F_{m-m_0}^{-1}(1-\gamma)$, где F_{m-m_0} – ф.р. с.в. $\chi_{m-m_0}^2$. Если до-

полнительно положить $\mathbf{W}_{1n}=(\mathbf{L}\mathbf{V}_B\mathbf{L}')^{-1}\mathbf{W}_{1n}=(\mathbf{L}\mathbf{V}_B\mathbf{L}')^{-1}$ и выбирать параметризацию альтернатив из условия $\mathbf{L}\mathbf{V}_B\mathbf{J}'=0$, то целевая функция в (8) также будет нормирована, и имеет место представление

$$\mathbf{B}'_n\mathbf{J}'\mathbf{W}_{2n}\mathbf{J}\mathbf{B}_n=\mathbf{B}'_n\mathbf{A}'\mathbf{W}_{2n}\mathbf{A}\mathbf{B}_n-\mathbf{B}'_n\mathbf{L}'\mathbf{W}_{2n}\mathbf{L}\mathbf{B}_n, \quad (10)$$

т.к. пространства, определяемые базисами $\mathbf{L}(\mathbf{V}_B^{1/2})'$ и $\mathbf{J}(\mathbf{V}_B^{1/2})'$ ортогональны, а слагаемые в (10) задают проекции вектора $(\mathbf{V}_B^{1/2})'\mathbf{B}_n$ на них и на их прямую сумму. Благодаря (10), можно выражать статистику критерия (9) через матрицы $\mathbf{A}$ и $\mathbf{L}$ (которые определяются содержательной постановкой задачи), минуя явное вычисление матрицы $\mathbf{J}$:

$$\omega_n=\operatorname{argmin}_{\omega_0\in\Omega}\zeta_n^2(\omega_0|\mathbf{L}),$$

$$\zeta_n^2(\omega_n|\mathbf{J})=\zeta_n^2(\omega_n|\mathbf{A})-\zeta_n^2(\omega_n|\mathbf{L})>t_\gamma, \quad (11)$$

где $\zeta_n^2(\omega_0|\mathbf{U})=\mathbf{B}'_n(\omega_0,0)\mathbf{U}'(\mathbf{U}\mathbf{V}_B\mathbf{U}')^{-1}\mathbf{U}\mathbf{B}_n(\omega_0,0)$.

В этом случае достигнутое на первом этапе значение целевой функции может быть использовано при вычислении статистики на втором этапе.

АСИМПТОТИЧЕСКИЕ СВОЙСТВА ИНДИКАТОРНЫХ СТАТИСТИК, ОЦЕНОК И КРИТЕРИЕВ ПРИ БЛИЗКИХ АЛЬТЕРНАТИВАХ

Изучим свойства процедуры (8)–(9) при близких альтернативах вида

$$H_n: \tau=\tau_n\in\{\tau: \|\tau\|<an^{-\beta}\}, \quad \beta>\frac{m}{2(m+1)}. \quad (12)$$

Сформулируем теорему о равномерной линейности индикаторных статистик, исследуем асимптотический уровень значимости индикаторного критерия и его асимптотическую мощность при близких альтернативах (12). Для этого без ограничения общности положим $\omega=0$ и будем ссылаться на следующие условия регулярности:

а) Элементы матриц $\mathbf{L}$ и $\mathbf{J}$ ограничены равномерно по объему выборки n .

б) $P\{\varepsilon_i<x\}=F_i(x|\mu_i)$, $F_i(x|\mu_i)\in F_{K,M,i}(\mathbf{c},\mathbf{D},\mathbf{p})$.

в) На альтернативных траекториях $F_{K,M,i}(\mathbf{c},\mathbf{D},\mathbf{p})$ ф.р. $F_i(x|\mu_i)$ непрерывно дифференцируемы в окрестностях точек $(x,\mu_i)=(c_k,0)$, $k=1,\dots,K-1$.

г) На альтернативных траекториях $F_{K,M,i}(\mathbf{c},\mathbf{D},\mathbf{p})$ плотности $f_i(x|\mu_i)$ удовлетворяют условию Липшица в окрестностях точек $(x,\mu_i)=(c_k,0)$, $k=1,\dots,K-1$ и величины $f_i(x|\mu_i)$ ограничены равномерно по n .

Введем функции отклика меток

$$\Psi_{1i}(u,\mu)=\sum_{k=1}^{K-1}[B_1(k)-B_1(k+1)][F_i(c_k|\mathbf{0})-F_i(c_k+u+\mathbf{d}'_k\mu|\mathbf{0})],$$

$$\Psi_{2i}(u,\mu)=\sum_{k=1}^{K-1}[B_2(k)-B_2(k+1)][F_i(c_k|\mathbf{0})-F_i(c_k+u+\mathbf{d}'_k\mu|\mathbf{0})],$$

$$\Psi_i(u,\mu)=u\Psi_{1i}(u,\mu)+\mu'\Psi_{2i}(u,\mu)$$

и для функций $\Psi_i(\omega)=\Psi_i(\mathbf{L}'_{1i}\omega,\mathbf{L}'_{2i}\omega)$, заданных на множестве $\Omega_i=\{\omega:\mathbf{L}'_{2i}\omega\in\mathbf{M}_{K,M}(\mathbf{c},\mathbf{D})\}$, будем рассматривать следующие условия регулярности:

д) При всех $\omega\in\Omega_i$ выполняется $\Psi_i(\omega)\leq 0$, и равенство здесь достигается, если и только если $|\mathbf{L}'_{1i}\omega|^2+|\mathbf{L}'_{2i}\omega|^2=0$.

е) Существуют $L_0, \delta_0>0$ такие, что при всех $\omega\in\Omega_i$, для которых $|\mathbf{L}'_{1i}\omega|^2+|\mathbf{L}'_{2i}\omega|^2<\delta_0$, выполняется $\Psi_i(\omega)\leq -L_0(|\mathbf{L}'_{1i}\omega|^2+|\mathbf{L}'_{2i}\omega|^2)$.

ж) Для любого $R>0$ существует $A(R)$ такое, что для всех $\omega\in\Omega_i$ при $|\mathbf{L}'_{1i}\omega|^2+|\mathbf{L}'_{2i}\omega|^2>R$ выполняется $\Psi_i(\omega)\leq -A(R)\sqrt{|\mathbf{L}'_{1i}\omega|^2+|\mathbf{L}'_{2i}\omega|^2}$.

Теорема 1 (о равномерной линейности индикаторных статистик). Пусть выполнены условия (а)–(г) и пусть последовательность $\{g_i; i\leq 1\}$ ограничена, а набор меток $\{B(1),\dots,B(K)\}$ удовлетворяет условию $\sum_{k=1}^K B(k)p_k=0$. Тогда каковы бы ни были траектории альтернативных распределений $F_{K,M,i}(\mathbf{c},\mathbf{D},\mathbf{p})$, если $\mu_i=\mathbf{J}'_{2i}\tau$, то для случайной функции

$$U_n(\omega,\tau)=\frac{1}{\sqrt{n}}\sum_{i=1}^n g_i\{B(s_i(\omega,0))-B(s_i(0,\tau))-(\mathbf{L}_{1i}\Psi_{1i}+\mathbf{L}_{2i}\Psi_{2i})'\omega+(\mathbf{J}_{1i}\Psi_{1i}+\mathbf{J}_{2i}\Psi_{2i})'\tau\}$$

при любых $\varepsilon>0, a>0, \gamma\in(1,\frac{1}{2(m+1)})$ выполняется

$$\lim_{n\rightarrow\infty}\sup_{\tau\in W_{2n}(an^\gamma)}P\left\{\sup_{\omega\in W_{2n}(an^\gamma)}|U_n(\omega,\tau)|\right\}=0,$$

где $W_{1n}(A)=\{\omega\in R^{m_0}: \|\omega\|\leq A\sqrt{n}, \mathbf{L}'_{2i}\omega\in\mathbf{M}_{K,M}(\mathbf{c},\mathbf{D})\forall i\}$,

$$W_{2n}(A)=\{\tau\in R^{m-m_0}: \|\tau\|\leq A\sqrt{n}, \mathbf{J}'_{2i}\tau\in\mathbf{M}_{K,M}(\mathbf{c},\mathbf{D})\forall i\},$$

$$\Psi_{1i}=\sum_{k=1}^{K-1}[B(k)-B(k+1)]f_i(c_k|\mathbf{0}),$$

$$\Psi_{2i}=\sum_{k=1}^{K-1}\mathbf{d}_k[B(k)-B(k+1)]f_i(c_k|\mathbf{0}).$$

Если теперь ввести матрицу чувствительности меток

$$\Psi_n=\frac{1}{n}\begin{pmatrix} \operatorname{diag}[\Psi_{111},\dots,\Psi_{11n}] & \mathbf{e}_1\Psi'_{121} & \mathbf{e}_n\Psi'_{12n} \\ \dots & \Psi_{211}\mathbf{e}'_1 & \dots \\ \Psi_{21n}\mathbf{e}'_n & \dots & 0 \\ & & & \Psi_{22n} \end{pmatrix},$$

$$\text{где } \Psi_{11i}=\sum_{k=1}^{K-1}[B_1(k)-B_1(k+1)]f_i(c_k|\mathbf{0}),$$

$$\Psi_{12i}=\sum_{k=1}^{K-1}[B_1(k)-B_1(k+1)]\mathbf{d}_k f_i(c_k|\mathbf{0}),$$

$$\Psi_{21i}=\sum_{k=1}^{K-1}[B_2(k)-B_2(k+1)]f_i(c_k|\mathbf{0}),$$

$$\Psi_{22i}=\sum_{k=1}^{K-1}[B_2(k)-B_2(k+1)]\mathbf{d}'_k f_i(c_k|\mathbf{0}),$$

то на основе теоремы 1 можно получить предельные разложения для индикаторных статистик из (8) и (9) при близких альтернативах τ_n вида (12).

Теорема 2. Пусть выполнены условия (а)–(г), существуют предельные матрицы

$$\Sigma_0=\lim_{n\rightarrow\infty}\mathbf{J}\Psi_n\mathbf{L}', \quad \Sigma_1=\lim_{n\rightarrow\infty}\mathbf{J}\Psi_n\mathbf{J}', \quad \Sigma_2=\lim_{n\rightarrow\infty}\mathbf{J}\mathbf{V}_B\mathbf{J}'>0$$

и ω_n – некоторая оценка параметров ω , для которой $n^{\beta_0}\|\omega_n\|=o_p(1)$ при каком-либо $\beta_0\in(\frac{m}{2(m+1)},\beta)$. Тогда

$$\mathbf{J}\mathbf{B}_n(\omega_n,0)=\mathbf{J}\mathbf{B}_n(0,\tau_n)+\sqrt{n}(\Sigma_0\omega_n-\Sigma_1\tau_n)+o_p(1).$$

Теорема 3. Пусть выполнены условия (а)–(ж) и существуют предельные матрицы

$$\Sigma_{01}=\lim_{n\rightarrow\infty}\mathbf{L}\Psi_n\mathbf{J}', \quad \Sigma_{11}=\lim_{n\rightarrow\infty}\mathbf{L}\Psi_n\mathbf{L}'<0, \quad \Sigma_{21}=\lim_{n\rightarrow\infty}\mathbf{L}\mathbf{V}_B\mathbf{L}'>0.$$

Тогда для индикаторной оценки (8) выполняется

$$\sqrt{n}\omega_n=-\Sigma_{11}^{-1}[\mathbf{L}\mathbf{B}_n(0,\tau_n)-\sqrt{n}\Sigma_{01}\tau_n]+o_p(1).$$

Теорема 2 позволяет судить о локальном поведении функции мощности индикаторного критерия (9), если

рассматривать близкие альтернативы с разной скоростью сходимости. При этом критерий (9) будет иметь асимптотический уровень значимости γ , если $\Sigma_0 = \mathbf{0}$. Для использования критерия (11) матрица $\mathbf{J}$ должна выбираться так, чтобы $\mathbf{L}\mathbf{V}_B\mathbf{J}' = \mathbf{0}$. По теореме 3 индикаторная оценка свободных параметров при близких альтернативах асимптотически нормальна, $\sqrt{n}$ – состоятельна и удовлетворяет условию теоремы 2, если $\Sigma_{01} = \mathbf{0}$ или $\tau_n = o(n^{-1/2})$.

Эти дополнительные требования будут выполнены, когда, например, матрицы $\mathbf{V}_B$ и Ψ_n пропорциональны или когда из $\mathbf{L}\mathbf{V}_B\mathbf{J}' = \mathbf{0}$ следует, что $\mathbf{J}\Psi_n\mathbf{L}' = \mathbf{0}$ и $\mathbf{L}\Psi_B\mathbf{J}' = \mathbf{0}$. Для проверки этого условия не обязательно строить матрицу $\mathbf{J}$. Достаточно убедиться в пропорциональности матриц $\mathbf{J}_0\Psi_n\mathbf{L}'$ и $\mathbf{J}_0\mathbf{V}_B\mathbf{L}'$, где $\mathbf{J}_0$ – любой базис, дополняющий $\mathbf{L}_0$ до $\mathbf{L}$, например, когда $\mathbf{A} = (\mathbf{L}', \mathbf{J}_0')$. При построении тестов будем проверять это условие их применимости.

ПРИМЕРЫ МЕТОК

Пример 1. Метки могут быть взяты в виде $\{B_1(k)\} = \{-1, 0, 1\}$, $\{B_2(k)\} = \{1, -Q, 1\}$, где $Q = 2p/(1-2p)$. Эти наборы пропорциональны наборам $\{R_i(k)\}$, если для плотности выполняется условие квантильной симметрии $f_i(-c) = f_i(c)$. Для моментов меток имеем $d_1^2 = 2p$, $\mathbf{D}_2 = Q$, $\mathbf{C} = \mathbf{0}$, так что матрица $\mathbf{V}_B$ будет диагональной. При условии квантильной симметрии имеем $\psi_{11i} = 2f_i(c)$, $\psi_{12i} = \psi_{21i} = 0$, $\psi_{22i} = 2f_i(c)(Q+1)$ и матрица Ψ_n тоже будет диагональной.

Пример 2. Здесь метки можно взять в виде $\{B_1(k)\} = \{-Q, -\alpha_A, \alpha_A, Q\}$, $\{B_2(k)\} = \{Q, -1, -1, Q\}$, где $Q = 2p/(1-2p)$. Пропорциональность с наборами $\{R_i(k)\}$ будет иметь место, если $f_i(-c) = f_i(c)$ и $\alpha_A = \alpha_i$, где $\alpha_i = (f_i(0) - f_i(c))/f_i(c)$. Величину α_A , таким образом, можно трактовать как априорную догадку о квантильной островершинности (или затянутости) гипотетического распределения. При отсутствии других соображений, рекомендуется брать $\alpha_A = Q/2$. Моменты меток имеют вид $d_1^2 = 2p(\alpha_A^2 + Q)$, $\mathbf{D}_2 = Q$, $\mathbf{C} = \mathbf{0}$, и вновь матрица $\mathbf{V}_B$ является диагональной. В условиях квантильной симметрии при дополнительном естественном требовании $f_i(0) = 0$ выполняется $\psi_{12i} = \psi_{21i} = 0$, $\psi_{11i} = 2f_i(c)(\alpha_A\alpha_i + Q)$, $\psi_{22i} = 2f_i(c)(Q+1)$, матрица Ψ_n диагональная.

Пример 3. Для квантильной регрессии метки $\{B_1(k)\} = \{-1/p, 1/(1-p)\}$ являются взвешенными знаками остатков. Эти метки всегда пропорциональны набору $\{R_1(k)\}$ и для них $d_1^2 = [p(1-p)]^{-1}$, $\mathbf{V}_B = d_1^2 \mathbf{I}_n$. При этом, если $f_i(0) = f(0)$, то $\Psi_n = \mathbf{I}_n f(0)/n$, и имеется пропорциональность матриц $\mathbf{V}_B$ и Ψ_n .

Пример 4. Для частного случая медианной регрессии метки $\{B_1(k)\} = \{-1, 1\}$ соответствуют знакам остатков. Здесь $d_1^2 = 1$, $\mathbf{V}_B = \mathbf{I}_n$, $\Psi_n = \mathbf{I}_n f(0)/n$ при $f_i(0) = f(0)$.

Если ф.р. F_i непрерывно дифференцируемы, то условия регулярности (д)–(ж) выполняются в примерах 3 и 4 всегда, в примере 1 – для всех симметричных ф.р. (и, по крайней мере, для близких к симметричным

ф.р.), а в примере 2 – для достаточно широкого класса симметричных и близких к симметричным распределений (равномерному, Лапласа, распределениям семейства Стюдента, в том числе Коши, и др.).

ЗАДАЧИ ОДНОФАКТОРНОГО АНАЛИЗА

Рассмотрим модель наблюдений

$$y_{ij} = \theta_j + \varepsilon_{ij}, j=1, \dots, q; i=1, \dots, n_j; n = \sum_{j=1}^q n_j, \quad (13)$$

где $\theta_1, \dots, \theta_q$ – неизвестные параметры.

Гипотеза об однородности q выборок по параметрам сдвига формулируется в виде

$$H_0: \theta_1 = \dots = \theta_q. \quad (14)$$

В качестве альтернативы выступает нарушение любого из этих равенств.

В примерах 3 и 4 при гипотезе свободен только параметр $\omega = \theta$ – общее значение параметра положения выборок. Запишем участвующие в процедуре (11) статистики:

$$\varsigma_n^2(\omega|\mathbf{L}) = \frac{1}{d_1^2 n} \left(\sum_{j=1}^q \sum_{i=1}^{n_j} B_{1ij}(\omega) \right)^2, \quad (15)$$

$$\varsigma_n^2(\omega|\mathbf{J}) = \sum_{j=1}^q \frac{n_j}{d_1^2} \left(\frac{1}{n_j} \sum_{i=1}^{n_j} B_{1ij}(\omega) - \frac{1}{n} \sum_{j=1}^q \sum_{i=1}^{n_j} B_{1ij}(\omega) \right)^2, \quad (16)$$

где $B_{ij}(\omega) = B_1(s(y_{ij} - \theta))$, $s(u) = (3 + \text{sign}(u))/2$.

В примерах 1 и 2 если все выборки имеют один неизвестный квантильный масштаб μ , то с точностью до обозначений $B_{kij}(\omega) = B_k(s(y_{ij} - \theta, \mu))$, $\omega = (\theta, \mu)'$ статистика теста $\varsigma_n^2(\omega|\mathbf{J})$ будет иметь вид (16). Здесь примеры 1 и 2 отличаются наборами меток и индикаторными функциями, которые были введены ранее. Статистика, используемая в (11) на первом этапе, отличается от (15):

$$\varsigma_n^2(\omega|\mathbf{L}) = \frac{1}{d_1^2 n} \left(\sum_{j=1}^q \sum_{i=1}^{n_j} B_{1ij}(\omega) \right)^2 + \frac{1}{d_2^2 n} \left(\sum_{j=1}^q \sum_{i=1}^{n_j} B_{2ij}(\omega) \right)^2. \quad (17)$$

Если выборки имеют разные (и неизвестные) квантильные масштабы μ_j , то условия применимости нормированного теста не выполняются. Возможен еще один вариант, когда общий параметр масштаба μ известен. В этом случае $\omega = \theta$, а статистики в (11) имеют вид (15) и (16) с точностью до обозначения $B_{ij}(\omega) = B_1(s(y_{ij} - \theta, \mu))$.

Перейдем к проверке гипотезы однородности q выборок по квантильным масштабам

$$H_0: \mu_1 = \dots = \mu_q. \quad (18)$$

Она может проверяться в рамках примеров 1 и 2 при известных и неизвестных параметрах положения выборок θ_j . Если известно, что $\theta_j = \theta$, но само значение θ неизвестно, то $\omega = (\theta, \mu)'$ и, как ранее, $B_{kij}(\omega) = B_k(s(y_{ij} - \theta, \mu))$, а статистики в (11) имеют вид (17) и

$$\varsigma_n^2(\omega|\mathbf{J}) = \sum_{j=1}^q \frac{n_j}{d_2^2} \left(\frac{1}{n_j} \sum_{i=1}^{n_j} B_{2ij}(\omega) - \frac{1}{n} \sum_{j=1}^q \sum_{i=1}^{n_j} B_{2ij}(\omega) \right)^2. \quad (19)$$

Если известно, что $\theta_j = \theta$ и параметр θ известен ($\theta = 0$ без ограничения общности), то

$$\varsigma_n^2(\omega|\mathbf{L}) = \frac{1}{d_2^2 n} \left(\sum_{j=1}^q \sum_{i=1}^{n_j} B_{2ij}(\omega) \right)^2, \quad (20)$$

а $\varsigma_n^2(\omega|\mathbf{J})$ дается (19) с точностью до обозначений $\omega = \mu$ и $B_{2ij}(\omega) = B_2(s(y_{ij}, \mu))$. Для случая, когда θ_j не-

известны и могут различаться, имеем $\omega = (\theta_1, \dots, \theta_q, \mu)'$, $B_{kij}(\omega) = B_k(s(y_{ij} - \theta, \mu))$, а статистики из (11) с точностью до этих обозначений даются формулами (17) и (19).

Наконец, для гипотезы однородности по сдвигам и масштабам

$$H_0: \theta_1 = \dots = \theta_q, \mu_1 = \dots = \mu_q, \quad (21)$$

в обозначениях $B_{kij}(\omega) = B_k(s(y_{ij} - \theta, \mu))$ и $\omega = (\theta, \mu)'$

целевая функция $\zeta_n^2(\omega|\mathbf{L})$ имеет вид (17), а выражение для статистики теста (11) сводится к

$$\zeta_n^2(\omega|\mathbf{J}) = \sum_{j=1}^q \frac{n_j}{d_1^2} \left(\frac{1}{n_j} \sum_{i=1}^{n_j} B_{1ij}(\omega) - \frac{1}{n} \sum_{j=1}^q \sum_{i=1}^{n_j} B_{1ij}(\omega) \right)^2 + \quad (22)$$

$$+ \sum_{j=1}^q \frac{n_j}{d_2^2} \left(\frac{1}{n_j} \sum_{i=1}^{n_j} B_{2ij}(\omega) - \frac{1}{n} \sum_{j=1}^q \sum_{i=1}^{n_j} B_{2ij}(\omega) \right)^2.$$

ДВУХФАКТОРНЫЙ АНАЛИЗ АДДИТИВНОЙ СХЕМЫ ДЕЙСТВИЯ ФАКТОРОВ

Рассмотрим модель эксперимента, которая не предполагает совместного влияния факторов:

$$y_{ijl} = a_j + b_l + \varepsilon_{ijl}, j=1, \dots, J; l=1, \dots, L; i=1, \dots, n_{jl}, \quad (23)$$

где $a_1, \dots, a_J, b_1, \dots, b_{L-1}$ – неизвестные параметры, $b_L=0$; ε_{ijl} – независимые случайные отклонения, информация о распределении которых сводится к модели одного из четырех примеров, рассмотренных ранее. Введем обозначения для объемов выборок $n_{j*} = \sum_{l=1}^L n_{jl}$, $n_{*l} = \sum_{j=1}^J n_{jl}$, $n_{**} = \sum_{l=1}^L n_{*l}$. Имея в виду метки остатков $B_{kijl}(\omega)$, будем использовать обозначения вида

$$B_{k*jl}(\omega) = \sum_{i=1}^{n_{jl}} B_{kijl}(\omega), \quad B_{k**l}(\omega) = \sum_{j=1}^J B_{k*jl}(\omega),$$

$$B_{k*j*}(\omega) = \sum_{l=1}^L B_{k*jl}(\omega), \quad B_{k***}(\omega) = \sum_{l=1}^L B_{k**l}(\omega).$$

Начнем с проверки гипотезы об отсутствии действия одного из факторов:

$$H_0: b_1 = \dots = b_{L-1} = 0. \quad (24)$$

Для примеров 3 и 4 $B_{kijl}(\omega) = B_k(s(y_{ijl} - a_j))$, где $\omega = (a_1, \dots, a_J)'$, $s(u) = (3 + \text{sign}(u))/2$. При этом

$$\zeta_n^2(\omega|\mathbf{L}) = \sum_{j=1}^J \frac{B_{1*j*}^2(\omega)}{d_1^2 n_{j*}}, \quad \zeta_n^2(\omega|\mathbf{J}) = \frac{\zeta'(\omega)\mathbf{Q}^{-1}\zeta(\omega)}{d_1^2}, \quad (25)$$

где поэлементно $[\zeta(\omega)]_l = B_{1**l}(\omega) - \sum_{j=1}^J n_{jl} \frac{B_{1*j*}(\omega)}{n_{j*}}$,

$[\mathbf{Q}]_{kl} = n_{*l} \delta_{kl} - \sum_{j=1}^J n_{jl} n_{jk} / n_{j*}$. При одинаковых объемах выборок, если $n_{jl} = n_{**}/(JL)$, то (25) упрощается:

$$\zeta_n^2(\omega|\mathbf{J}) = \sum_{l=1}^L \frac{B_{1**l}^2(\omega)}{d_1^2 n_{*l}} - \frac{B_{1***}^2(\omega)}{d_1^2 n_{**}}.$$

Для примеров 1 и 2 при известных и одинаковых для всех выборок масштабах шумов μ можно использовать точно такие же статистики вида (25), но при этом полагая $B_{kijl}(\omega) = B_k(s(y_{ijl} - a_j, \mu))$. В случае, если общий параметр масштаба μ неизвестен, имеем $\omega = (a_1, \dots, a_J, \mu)'$, $B_{kijl}(\omega) = B_k(s(y_{ijl} - a_j, \mu))$, а в (25) изменится вид статистики

$$\zeta_n^2(\omega|\mathbf{L}) = \sum_{j=1}^J \frac{B_{1*j*}^2(\omega)}{d_1^2 n_{j*}} + \frac{B_{2***}^2(\omega)}{d_2^2 n_{**}},$$

при этом вид $\zeta_n^2(\omega|\mathbf{J})$ останется прежним. Наконец, если допустить, что масштабы выборок могут различаться, то нормированный тест построить не удастся, т.к. условия его применимости не выполняются.

Перейдем к проверке гипотезы об отсутствии действия обоих факторов:

$$H_0: a_1 = \dots = a_J, b_1 = \dots = b_{L-1} = 0. \quad (26)$$

Здесь в примерах 3 и 4, а также в примерах 1 и 2 при известных масштабах шумов имеем $\omega = a$,

$$\zeta_n^2(\omega|\mathbf{L}) = B_{1***}^2(\omega) / (d_1^2 n_{**}), \quad (27)$$

$$\zeta_n^2(\omega|\mathbf{J}) = \sum_{j=1}^J \frac{B_{1*j*}^2(\omega)}{d_1^2 n_{j*}} - \frac{B_{1***}^2(\omega)}{d_1^2 n_{**}} + \frac{\zeta'(\omega)\mathbf{Q}^{-1}\zeta(\omega)}{d_1^2}, \quad (28)$$

примеры отличаются индикаторами и метками: для примеров 1 и 2 $B_{kijl}(\omega) = B_k(s(y_{ijl} - a, \mu))$, где μ – известная величина, в то время как для примеров 3 и 4 $B_{kijl}(\omega) = B_k(s(y_{ijl} - a))$. При одинаковых объемах выборок n_{jl} формула (28) упрощается, подобно (25). Как и ранее, допущение разных масштабов шумов в выборках приводит к невозможности построения нормированного теста.

При неизвестном (но общем для всех выборок) масштабе шумов μ в рамках примеров 1 и 2 имеем $\omega = (a, \mu)'$, $B_{kijl}(\omega) = B_k(s(y_{ijl} - a, \mu))$, тестовая статистика для проверки гипотезы (26) будет иметь вид (28), но для оценки свободных параметров вместо (27) будет использоваться целевая функция

$$\zeta_n^2(\omega|\mathbf{L}) = \frac{B_{1***}^2(\omega)}{d_1^2 n_{**}} + \frac{B_{2***}^2(\omega)}{d_2^2 n_{**}}. \quad (29)$$

Следующий вид гипотезы включает предположение об отсутствии действия факторов на масштабы шумов выборок в рамках примеров 1 и 2:

$$H_0: \mu_{11} = \dots = \mu_{JL}. \quad (30)$$

Здесь $\omega = (a_1, \dots, a_J, b_1, \dots, b_{L-1}, \mu)'$ – вектор свободных параметров, $B_{kijl}(\omega) = B_k(s(y_{ijl} - a_j - b_l, \mu))$,

$$\zeta_n^2(\omega|\mathbf{L}) = \sum_{j=1}^J \frac{B_{1*j*}^2(\omega)}{d_1^2 n_{j*}} + \frac{\zeta'(\omega)\mathbf{Q}^{-1}\zeta(\omega)}{d_1^2} + \sum_{j=1}^J \sum_{l=1}^L \frac{B_{2*jl}^2(\omega)}{d_2^2 n_{jl}},$$

$$\zeta_n^2(\omega|\mathbf{J}) = \sum_{j=1}^J \sum_{l=1}^L \frac{B_{2*jl}^2(\omega)}{d_2^2 n_{jl}} - \frac{B_{2***}^2(\omega)}{d_2^2 n_{**}}. \quad (31)$$

Для аналогичной гипотезы вида

$$H_0: b_1 = \dots = b_{L-1} = 0, \mu_{11} = \dots = \mu_{JL} \quad (32)$$

имеем $\omega = (a_1, \dots, a_J, \mu)'$, $B_{kijl}(\omega) = B_k(s(y_{ijl} - a_j, \mu))$,

$$\zeta_n^2(\omega|\mathbf{L}) = \sum_{j=1}^J \frac{B_{1*j*}^2(\omega)}{d_1^2 n_{j*}} + \sum_{j=1}^J \sum_{l=1}^L \frac{B_{2*jl}^2(\omega)}{d_2^2 n_{jl}}, \quad (33)$$

структура $\zeta_n^2(\omega|\mathbf{J})$ будет даваться (с точностью до обозначений) суммой статистик вида (25) и (31).

Наконец, для проверки гипотезы

$$H_0: a_1 = \dots = a_J, b_1 = \dots = b_{L-1} = 0, \mu_{11} = \dots = \mu_{JL} \quad (34)$$

получаем $\omega = (a, \mu)'$, $B_{kijl}(\omega) = B_k(s(y_{ijl} - a, \mu))$, целевая функция $\zeta_n^2(\omega|\mathbf{L})$ имеет вид (33), но $\zeta_n^2(\omega|\mathbf{J})$ есть сумма статистик вида (28) и (31).

Тесты, полученные здесь для проверки гипотез (32) и (34), можно применять только в условиях примера 1. Для примера 2 условия применимости нормированного теста будут выполнены, только если $\alpha_A = \alpha$.

ЗАКЛЮЧЕНИЕ

Пользуясь (11), можно построить тесты для проверки линейных гипотез в двухфакторной модели, которая учитывает совместное влияние факторов, но эти задачи не являются темой данной работы. Это же можно сказать и о проверке гипотез об отсутствии влияния одного из факторов на масштабы шумов.

Целевые функции $\zeta_n^2(\omega|\mathbf{L})$, которые используются для оценки свободных параметров, являются кусочно-постоянными, и это требует специальных методов минимизации, которые здесь не рассмотрены.

Согласно теоремам 2 и 3, при гипотезе предельным распределением статистик $\zeta_n^2(\omega|\mathbf{J})$ будет хи-квадрат с соответствующим числом степеней свободы. Однако при уме-

ренных объемах выборки для выбора порога тестов можно использовать процентные точки распределения статистик $\zeta_n^2(\omega|\mathbf{J})$, в которых величины $B_{kjl}(\omega)$ заменены на $B_{kjl}(\zeta_{ijl})$, где ζ_{ijl} – независимые дискретные с.в., принимающие значения $1, \dots, K$ с вероятностями $p_1, \dots, p_K$.

Обращает на себя внимание структура тестовых статистик $\zeta_n^2(\omega|\mathbf{J})$. Она во многом аналогична числителю отношения Фишера–Снедекора, которое используется для проверки линейных гипотез в гауссовском дисперсионном анализе. В то же время, благодаря описанию масштаба распределений шумов с помощью интерквантильного размаха, реализуется возможность проверки выборок на однородность не только по сдвигам, но и по масштабам.

ЛИТЕРАТУРА

1. Koenker R., Bassett G. Regression quantiles // *Econometrica*. 1978. Vol. 46. P. 33–50.
2. Koenker R., Bassett G. Tests for linear hypothesis and L_1 estimation // *Econometrica*. 1982. Vol. 50. P. 1577–1584.
3. Koenker R., Portnoy S. M estimation of multivariate regressions // *JASA*. 1990. Vol. 85. Issue 412. P. 1060–1068.
4. Болдин М.В., Симонова Г.И., Тюрин Ю.Н. Знаковый статистический анализ линейных моделей. М.: Наука, 1997.
5. Тарасенко П.Ф. Оптимальные тесты, основанные на индикаторах событий // Вестник Томского гос. ун-та. Прил. № 1(1), сент.2002. Докл. IV Всеросс. конф. «Нов. информ. технологии в исслед. сложн. структур». Томск, 10–13 сентября 2002. С. 185–190.
6. Тарасенко П.Ф. Проверка гипотез о параметрах линейной модели на основе индикаторных признаков // Обозрение прикладной и промышленной математики. 2003. Т. 10. Вып. 2. IV Всеросс. симп. по прикл. и пром. матем., осенняя сессия: Тез. докл. Ч. I, С. 515–517.

Статья представлена кафедрой теоретической кибернетики факультета прикладной математики и кибернетики Томского государственного университета, поступила в научную редакцию «Кибернетика» 20 апреля 2004 г.