УДК 519.233 : 681.3.01

Г. И. ПЕРЕТЯГИН

(Новосибирск)

ОТБОР ВЫДЕЛЯЮЩИХСЯ НАБЛЮДЕНИЙ И КРИТЕРИИ СДВИГА

При автоматизации экспериментальных исследований на стадии предварительной обработки поступающей информации часто возникает задача обнаружения элементов, выделяющихся из однородного (в некотором смысле) множества. В частности, к таковым относятся, например, задача отбора «информативных» фрагментов при обработке изображений, обнаружение «сигнальных» точек в статистически однородных случайных полях, отбраковка аномальных измерений и т. п.

1. В своей простейшей формулировке эти проблемы можно свести к следующему виду. В некоторой последовательности наблюдений, состоящей из n измеренных величин $(x_1, x_2, \ldots, x_n)$, имеется одно или несколько значений, не удовлетворяющих модели исследуемого процесса. Требуется обнаружить эти величины. Основная трудность здесь состоит в том, что измерения сопровождаются неустранимым шумом $(x_i = x_{0i} + \xi_i)$ и проверка наблюдаемых величин должна основываться на анализе статистических характеристик процесса. Нужно отметить, что если проблема выделения одного выпадающего наблюдения допускает общее решение с точки зрения теории проверки гипотез [1], то обнаружение нескольких выделяющихся элементов (что более существенно для практики) до сих пор представляет собой «камень преткновения». Это связано с тем, что применяемые обычно последовательные процедуры отбора выпадающих наблюдений (начиная с экстремального) с теоретической точки зрения не вполне удовлетворительны — здесь принципиально нельзя учесть влияние всех выбросов на распределение критических статистик. Неудивительно, что эти статистики иногда искажаются выпадающими наблюдениями настолько, что выходят из своей критической зоны (подробнее об эффекте «маскирования» см. в работах [2, 3]). Очевидный путь повышения устойчивости критериев отбора выпадающих наблюдений состоит в том, чтобы основывать их на «центральной» части выборки. Цель данной работы и состоит в реализации такого подхода. Прежде всего находятся оптимальные решающие правила для фиксированного числа выделяющихся наблюдений. Принципы инвариантности и достаточности в применении к данным решающим правилам позволяют непосредственно найти оптимальные критерии сдвига в задачах с «мешающими» параметрами. Наконец, при неизвестном общем числе выпадающих наблюдений приводится оптимальная процедура отбора этих наблюдений.

2. Пусть $X_1, X_2, \ldots, X_n$ — независимые одинаково распределенные случайные величины, имеющие общую функцию распределения $F_\theta(X)$, где $\theta = \theta_0 \in \Theta$. Объем выборки n фиксирован. Альтернатива состоит в том, что некоторые из случайных величин $X_{\nu_1}, X_{\nu_2}, \ldots, X_{\nu_k}$ имеют ту же самую общую функцию распределения $F_\theta(X)$, но $\theta_{\nu_j} \neq \theta_0$, $j = \overline{1, k}$. Будем считать, что заранее нельзя отдать предпочтение ни одному из C_n^k возможных наборов $(X_{\nu_1}, \ldots, X_{\nu_k})$, удовлетворяющих альтернативному требованию задачи. В связи с этим естественно ограничиться лишь симметричными решающими правилами. Для этого сведем задачу к проверке многих гипотез $H_0, H_1, \ldots, H_M$, $M = C_n^k$, и потребуем, чтобы гипотезы H_i, $i = \overline{1, M}$, были симметричными.

Обозначим через $f(X, \theta)$ плотность распределения случайной величины $X=(X_1, X_2, \ldots, X_n)$, когда параметр равен θ. Кроме того, пусть $(\lambda_1^\nu, \lambda_2^\nu)$ обозначают всевозможные разбиения $(\nu=\overline{1, M})$ множества

$$\prod_{i=1}^{n}$$

При альтернативных гипотезах $H_i^{(\lambda_1^i, \lambda_2^i)}$ $(i=\overline{1, M})$ плотность совместного распределения компонент вектора X имеет вид

$$h_i(X) = h(X|S_i) = \prod_{j=k+1}^{n} f(X_{ij}, \Theta_0) \prod_{l=1}^{k} f(X_{il}, \Theta_{il}) =$$

$$= \prod_{j=1}^{n} f(X_j, \Theta_0) \prod_{l=1}^{k} \frac{f(X_{il}, \Theta_{il})}{f(X_{il}, \Theta_0)}. \qquad (1)$$

По данному выборочному значению $x=(x_1, x_2, \ldots, x_n)$ требуется отдать предпочтение одной из гипотез $H_0, H_1, \ldots, H_M$. Здесь нерандомизированное решающее правило может быть представлено измеримым разбиением выборочного пространства V_n на $(M+1)$ взаимно непересекающихся подмножеств $(\omega_0, \omega_1, \ldots, \omega_M)$, где $\bigcup_{i=0}^{M} \omega_i = V_n$ и $\omega_i \bigcap_{i \neq j} \omega_j = \varnothing$, так что выбирается i-я альтернативная гипотеза H_i, когда $x \in \omega_i$. Поступая обычным способом, введем критическую функцию $\varphi(x) = (\varphi_0(x), \varphi_1(x), \ldots, \varphi_M(x))$, где $\varphi_i(x)$ равна 1 в области ω_i и 0 вне этой области (при желании $\varphi(x)$ можно продолжить до рандомизированной критической функции [4]). Кроме того, пусть γ_ν обозначает перестановку компонент n-мерного вектора X, соответствующую разбиению $(\lambda_1^\nu, \lambda_2^\nu)$ множества $\{1, 2, \ldots, n\}$, и Γ — группа всех таких перестановок. Тогда требование симметрии будет означать, что решающая функция $\varphi(x)$ симметрична по Γ, т. е. замена $\gamma_\nu x$ на $\gamma_\mu x$ приводит к такому переупорядочению функций $\varphi_i(\gamma_\nu x)$, $i=\overline{1, M}$, что в целом $\varphi(x)$ остается неизменной: $\varphi(\gamma_\nu x) = \varphi(\gamma_\mu x)$, $\gamma_\nu, \gamma_\mu \in \Gamma$. Нетрудно заметить, что в этом случае вероятность принятия гипотезы H_i при условии, что она верна, $P(H_i|H_i)$ не зависит от i для $i=\overline{1, M}$. Принимая данные положения во внимание, будем рассматривать решающие правила $\varphi(x) = (\varphi_0(x), \varphi_1(x), \ldots, \varphi_M(x))$, максимизирующие вероятности правильной классификации — $P(H_i|H_i)$, $i=\overline{1, M}$ при условии, что $P(H_0|H_0) = 1-\alpha$, $0<\alpha<1$.

3. Сформулируем теперь байесовскую задачу решения о выборе одной из гипотез $H_0, H_1, \ldots, H_M$. При этом будем предполагать, что θ_0 известно. Прежде всего введем функцию $L_j(H_i)$, определяющую потери, связанные с принятием гипотезы H_j, когда верна гипотеза H_i. Примем ее равной

$$L_j(H_i) = 1 - \delta_{ij} = \begin{cases} 1, & i \neq j; \\ 0, & i = j. \end{cases} \qquad (2)$$

Тогда для данного правила множественного решения $\varphi(x)$ математическое ожидание потерь (функция риска) есть вектор $(R(H_0, \varphi), R(H_1, \varphi), \ldots, R(H_M, \varphi))$,

где

$$R(H_i, \varphi) = \int_{V_n} \sum_{j=0}^{M} L_j(H_i)\, \varphi_j(x)\, h(x|s_i)\, dx = 1 - E_{s_i}\varphi_i(x). \qquad (3)$$

Используя байесовский принцип, будем считать, что p_i— априорные вероятности гипотез H_i, а истинное распределение имеет плотность $h(x|s_i)$. В этом случае байесовский риск относительно $p=(p_0, p_1, \ldots, p_M)$ есть величина

$$r(p, \varphi) = \sum_{i=0}^{M} p_i R(H_i, \varphi) = 1 - \sum_{i=0}^{M} p_i E_{s_i}\varphi_i(x). \qquad (4)$$

Проблема выбора $\varphi(x)$, минимизирующей данный риск, эквивалентна проблеме выбора $\varphi(x)$ для максимизации вероятности правильной классификации [4]. Известное байесовское решение относительно данного p задается следующим правилом: принять гипотезу H_j, если

$$p_j h_j(x) = \max_i p_i h_i(x), \quad i=0, 1, \ldots, M. \qquad (5)$$

Другими словами, каждое правило $\varphi(x)$, для которого $\{\varphi_j(x)=1, \varphi_i(x)=0, j\neq i\}$, если $p_j h_j(x)$ больше любого другого $p_i h_i(x)$, имеет минимальный байесовский риск. (Здесь нужно отметить, что байесовское правило разрешает определенную рандомизацию среди тех j_l, для которых $p_{j_l} h_{j_l}(x) = \max_i p_i h_i(x)$, $l=\overline{1, m}$; следует лишь удовлетворить требованию $\sum_{l=1}^{m} \varphi_{j_l} = 1$.) Принимая во внимание симметрию задачи, будем брать априорные вероятности симметричными, исходя из того, что вероятность принятия гипотезы H_i не должна зависеть от перестановок наблюдений:

$$p_i = p(H_i) = p(H_j) = p_j (j=\overline{1, M}).$$

В этом случае симметричные априорные распределения должны давать равные веса всем гипотезам H_i, $i=1, M$. Поэтому приходим к представлению априорных распределений в виде $\hat{p}_0 = p(H_0) = 1 - Mp^0$, $\hat{p}_i = p(H_i) = p^0$ для всех $i=1, M$ и $0 \leqslant p^0 \leqslant 1/M$. Используя правило (5), найдем байесовскую стратегию для распределения $\hat{p}(H_i)$.

Пусть $(\lambda_1^\nu, \lambda_2^\nu)$ — произвольное разбиение множества $\{1, 2, \ldots, n\}$, а $(\lambda_1^\xi, \lambda_2^\xi)$ — такое разбиение, что как только $x_{\xi_i} = \gamma_\xi x_i$, то

$$\max_\nu \prod_{j=1}^{k} \frac{f_j(x_{\nu_j})}{f_0(x_{\nu_j})} = \prod_{j=1}^{k} \frac{f_j(x_{\xi_j})}{f_0(x_{\xi_j})} = W.$$

Тогда каждое решающее правило $\hat{\varphi}(x)$, для которого

$$\hat{\varphi}_0(x) = 1, \quad \text{если} \quad W < \frac{1 - Mp^0}{p^0};$$

$$\hat{\varphi}_\nu(x) = 1, \quad \text{если} \quad \prod_{i=1}^{k} \frac{f_i(x_{\nu_i})}{f_0(x_{\nu_i})} = W > \frac{1 - Mp^0}{p^0} \quad (\nu = \overline{1, M}), \qquad (6)$$

является байесовским относительно $\hat{p}(H_i)$ и максимизирует вероятность выбора правильного решения.

Если число выпадающих наблюдений (k) неизвестно, то можно использовать минимаксное решающее правило, соответствующее наи-

менее благоприятному априорному распределению $\hat{p}^*(H_i)$, для которого

$$\min_{\varphi} r(\hat{p}^*, \varphi) = \max_{\hat{p}} \min_{\varphi} r(\hat{p}, \varphi).$$

В этом случае

$$r(\hat{p}, \varphi) = 1 - \sum_{i=0}^{M} \hat{p}_i E_{s_i}\varphi_i(x) = 1 - (1 - Mp^0) E_{s_0}\varphi_0(x) - p^0 \sum_{i=1}^{M} E_{s_i}\varphi_i(x),$$

где $0 < p^0 < 1/M$, $M = C_n^k$. Выбрав для определенности $p^0 = \varepsilon/M$, где $0 < \varepsilon < 1$, заметим, что $r(\hat{p}, \varphi)$ возрастает по переменной p^0 с ростом k, достигая своего максимального значения при $k = n/2$. Отсюда следует, что наименее благоприятное априорное распределение $\hat{p}^*$ соответствует либо $k = n/2$, либо некоторому $k_{\max} < n/2$. Отметим, что решающее правило (6) будет симметричным байесовским правилом и относительно наименее благоприятного априорного распределения.

Для перехода к классической постановке теории проверки гипотез нужно положить $\frac{1 - Mp^0}{p^0} = C_\alpha$; в этом случае риск $R(H_0, \varphi)$ называется уровнем значимости критерия $\alpha = 1 - E_{s_0}\varphi_0(x) = R(H_0, \varphi)$. Если определить ρ как число разбиений $(\lambda_1^\nu, \lambda_2^\nu)$, для которых $\prod_{j=1}^{k} \frac{f_j(x_{\nu j})}{f_0(x_{\nu j})} = W$, и найти C_α такое, что $E_{s_0}\varphi_0(x) = 1 - \alpha$, то из соотношения (6) непосредственно следует, что верна гипотеза H_0 при $W < C_\alpha$ и гипотезы H_ν при $\prod_{i=1}^{k} \frac{f_i(x_{\nu_i})}{f_0(x_{\nu_i})} = W > C_\alpha$:

$$\hat{\varphi}_0(x) = 1, \text{ если } W < C_\alpha;$$

$$\hat{\varphi}_j(x) = \frac{1}{\rho}, \text{ если } \prod_{i=1}^{k} \frac{f_i(x_{j_i})}{f_0(x_{j_i})} = \max_{\nu} \prod_{i=1}^{k} \frac{f_i(x_{\nu_i})}{f_0(x_{\nu_i})} = W > C_\alpha. \qquad (6')$$

(При $W = C_\alpha$ возможна рандомизация: $\varphi_0(x) = \eta$ и $\varphi_j(x) = (1 - \eta)/\rho$.) Данное правило является симметричным допустимым, в силу того что оно симметрично по Γ и байесовское относительно $\hat{p}(H_i)$, где $p^0 = 1/(C_\alpha + M)$. Более того, пусть $\tilde{\varphi}(x) = (\tilde{\varphi}_0(x), \tilde{\varphi}_1(x), \ldots, \tilde{\varphi}_M(x))$ есть некоторое другое симметричное решающее правило, имеющее $E_{s_0}\tilde{\varphi}_0(x) \geqslant E_{s_0}\hat{\varphi}_0(x)$. Тогда, поскольку правило (6′) байесовское относительно $\hat{p}(H_i)$, $r(\hat{p}, \tilde{\varphi}) \geqslant r(\hat{p}, \hat{\varphi})$ и $r(\hat{p}, \tilde{\varphi}) - r(\hat{p}, \hat{\varphi}) = \sum_{i=0}^{M} \hat{p}(H_i)\{E_{s_i}\hat{\varphi}_i(x) - E_{s_i}\tilde{\varphi}_i(x)\} \geqslant 0$. По предположению $E_{s_0}\tilde{\varphi}_0(x) \geqslant E_{s_0}\hat{\varphi}_0(x)$ и $\hat{p}(H_i) > 0$ получаем $\sum_{i=1}^{M} E_{s_i}\hat{\varphi}_i(x) \geqslant \sum_{i=1}^{M} E_{s_i}\tilde{\varphi}_i(x)$. Учитывая, что $\sum_{i=1}^{M} E_{s_i}\varphi_i(x)$ характеризует мощность критерия $\varphi(x)$, резюмируем вышеизложенное следующим образом.

В классе всех симметричных правил, удовлетворяющих условиям $P(H_0|H_0) = 1 - \alpha$ и $P(H_i|H_i) = P(H_j|H_j)$, $j = \overline{1, M}$, каждый критерий $\hat{\varphi}(x)$ вида (6′) для некоторого $C_\alpha (0 \leqslant \eta(x) \leqslant 1)$ является наиболее мощным объема α для проверки гипотезы H_0 против H_i, $i = \overline{1, M}$. (Отметим, что данное предложение — обобщение результата Фергюсона [1], относящегося к проверке одного выпадающего наблюдения.)

Если альтернативные гипотезы сложные, то для семейств распределений $h_i(x) = h(x|s_i)$ с монотонным отношением правдоподобия

данное предложение приводит к равномерно наиболее мощным (РНМ) симметричным критериям. Именно если существуют функции $T_\nu(x)$ такие, что отношение $h(x|s_\nu)/h(x|s_0)$ является неубывающей функцией от $T_\nu(x)$ и $\frac{h(x|s_\nu)}{h(x|s_\mu)} \gtrless 1$, когда соответственно $T_\nu(x)/T_\mu(x) \gtrless 1$, то правило (6′) сводится к следующему:

$$\hat{\varphi}_0(x) = 1 \text{ при } \max_\nu T_\nu(x) < C_\alpha;$$
$$\hat{\varphi}_j(x) = 1 \text{ при } T_j(x) = \max_\nu T_\nu(x) > C_\alpha. \tag{7}$$

Данное правило является РНМ симметричным.

4. Вообще говоря, распределение результатов наблюдений, получаемых в эксперименте, зависит дополнительно от некоторых «мешающих» параметров. В этом случае в качестве семейств с монотонным отношением правдоподобия обычно берут распределения, инвариантные относительно подходящим образом выбранной группы преобразований [5]. Для иллюстрации рассмотрим следующую задачу. Пусть $(x_1, \ldots, x_n)$ — независимые случайные переменные, относительно распределения которых можно выдвинуть $(C_n^k + 1)$ гипотез:

$$H_0 : x_j \sim \mathcal{N}(\mu, \sigma^2),\ j = \overline{1, n};$$
$$H_\nu : x_{\nu_i} \sim \mathcal{N}(\mu + \eta_i, \sigma^2),\ i = \overline{1, k}, \text{ и } x_{\nu_j} \sim \mathcal{N}(\mu, \sigma^2),\ j = \overline{k+1, n},$$

где μ, σ^2 — неизвестные параметры, а $(\eta_1, \eta_2, \ldots, \eta_k)$ — неизвестные положительные числа. Нетрудно заметить, что проблема инвариантна к изменениям параметров сдвига (μ) и масштаба (σ) в том смысле, что умножение всех x_i на постоянную λ и добавление к ним одинаковой константы оставляет каждую из гипотез неизменной. Естественно потребовать, чтобы решающее правило $\varphi(x)$ для выбора одной из $(M+1)$ гипотез было инвариантным относительно изменений параметров сдвига и масштаба:

$$\varphi_i(x_1, x_2, \ldots, x_n) = \varphi_i(\lambda x_1 + u, \lambda x_2 + u, \ldots, \lambda x_n + u)$$

для всех $i = 0, 1, \ldots, M$, u и λ. Для таких правил вероятности $P(H_i|H_i)$ не зависят от μ и σ^2.

Выделение класса решающих правил, инвариантных относительно изменений положения и масштаба наблюдений, означает выделение в исходном выборочном пространстве некоторого класса $\mathcal{A}_1$ инвариантных событий A таких, что для любых действительных u и λ имеет место эквивалентность [5]

$$\{(x_1, x_2, \ldots, x_n) \in A\} \Leftrightarrow \{(\lambda x_1 - u, \lambda x_2 - u, \ldots, \lambda x_n - u) \in A\}.$$

Если определить

$$\hat{f}(x_1, x_2, \ldots, x_n) = \int_0^\infty \int_{-\infty}^\infty f(\lambda x_1 - u, \lambda x_2 - u, \ldots, \lambda x_n - u)\lambda^{n-2}\, du d\lambda$$

и

$$\hat{f}_0(x_1, x_2, \ldots, x_n) = \int_0^\infty \int_{-\infty}^\infty f_0(\lambda x_1 - u, \lambda x_2 - u, \ldots, \lambda x_n - u)\lambda^{n-2}\, du d\lambda,$$

то

$$\frac{\hat{f}(x_1, x_2, \ldots, x_n)}{\hat{f}_0(x_1, x_2, \ldots, x_n)} = \frac{\hat{f}(\lambda x_1 - u, \lambda x_2 - u, \ldots, \lambda x_n - u)}{\hat{f}_0(\lambda x_1 - u, \lambda x_2 - u, \ldots, \lambda x_n - u)}$$

и для любого $A \subset \mathscr{A}_1$

$$P(A) = \int_A \frac{\widehat{f}(x)}{\widehat{f}_0(x)} f_0(x)\, dx.$$

В нашем примере имеем сложную нулевую гипотезу, состоящую из плотностей вида

$$h_0(x \mid \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i-\mu)^2 + \sum_{i=k+1}^{n}(x_{\nu i}-\mu^2)\right]\right\}.$$

Так как распределение инвариантной статистики критерия не зависит от параметров сдвига и масштаба, то для упрощения дальнейших вычислений положим $\mu=0$ и $\sigma^2=1$. Тогда

$$\widehat{h}_0(x \mid 0, 1) = \frac{1}{2} n^{-\frac{1}{2}} \pi^{-\frac{n}{2}+\frac{1}{2}} \Gamma\left(\frac{n}{2} - \frac{1}{2}\right)\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{-\frac{n}{2}+\frac{1}{2}}, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$\widehat{h}_\nu(x \mid \eta, 1) = n^{-\frac{1}{2}}(2\pi)^{-\frac{n}{2}+\frac{1}{2}}\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{-\frac{n}{2}+\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\sum_{i=1}^{k}\eta_i^2 - \frac{1}{2}\left(\sum_{i=1}^{k}\eta_i\right)^2\right)\right\}\int_0^\infty e^{-\frac{1}{2}(\lambda_1^2 - 2\lambda_1 t)}\lambda_1^{n-2}\, d\lambda_1,$$

$$t = \frac{\sum_{i=1}^{k} r_i (x_{\nu i} - \bar{x})}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{1/2}}.$$

Следовательно,

$$\frac{\widehat{h}_\nu(x \mid \eta, 1)}{\widehat{h}_0(x \mid 0, 1)} = C(\eta)\int_0^\infty \exp\left\{-\frac{1}{2}\lambda^2 + \lambda \frac{\sum_{i=1}^{k}\eta_i(x_{\nu i}-\bar{x})}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{1/2}}\right\}\lambda^{n-2} d\lambda.$$

Если $\eta_1=\eta_2=\ldots=\eta_k=\eta$, то полученное отношение будет монотонно возрастающей функцией от $T_\nu(x) = \dfrac{\sum_{i=1}^{k}(x_{\nu i}-\bar{x})}{\left[\sum_{1}^{n}(x_i-\bar{x})^2\right]^{1/2}}$. В соответствии с (7) равномерно наиболее мощным симметричным объема α решающим правилом, инвариантным относительно изменений параметров сдвига и масштаба, будет следующее:

$$\widehat{\varphi}_0(x) = \begin{cases} 1, & \text{если } \max\limits_{\nu} T_\nu(x) < C_\alpha; \\ 0, & \text{если } \max\limits_{\nu} T_\nu(x) \geqslant C_\alpha; \end{cases}$$

$$\widehat{\varphi}_j(x) = \begin{cases} 1, & \text{если } T_j(x) = \max\limits_{\nu} T_\nu(x) \geqslant C_\alpha; \\ 0 & \text{в остальных случаях.} \end{cases}$$

Константа C_α выбирается так, чтобы $P(H_0|H_0) = 1-\alpha$. Данное правило имеет максимальную величину $P(H_i|H_i)$ для $i \neq 0$ в классе всех симметричных правил независимо от $\eta > 0$. Можно заметить, что

$$\max_{\nu} T_\nu(x) = \max_{\nu} \frac{\sum\limits_{i=1}^{k} (x_{\nu_i} - \bar{x})}{S} = \frac{\sum\limits_{i=n-k+1}^{n} (x_{(i)} - \bar{x})}{S},$$

где $x_{(1)} < x_{(2)} < \ldots < x_{(n)}$ — порядковые статистики выборки. Кроме того, не существует РНМ симметричного инвариантного критерия, если $\eta_1 \neq \eta_2 \neq \ldots \neq \eta_k$.

5. Для поиска соответствующих тестов (при произвольных η_i) рассмотрим задачу с точки зрения общей линейной гипотезы. Именно будем считать, что n наблюдаемых величин $(x_1, \ldots, x_n)$ являются линейными комбинациями m неизвестных постоянных $\theta_1, \theta_2, \ldots, \theta_m$ и «ошибок» $\xi_1, \xi_2, \ldots, \xi_n$:

$$x_i = p_{i1}\theta_1 + p_{i2}\theta_2 + \ldots + p_{im}\theta_m + \xi_i, \quad i = \overline{1, n},$$

где $\{p_{ij}\} = P$ — матрица известных постоянных коэффициентов. Целью анализа является получение выводов относительно $\{\xi_i\}$: соответствуют ли они основным предположениям ($\xi_i \sim \mathcal{N}(0, \sigma^2)$, σ^2 неизвестно) либо необходимо «исключение» соответствующих x_{j_l} (при $\xi_{j_l} \sim \mathcal{N}(\eta_l, \sigma^2)$, $l = \overline{1, k}$).

Для представления проблемы в канонической форме введем выборочное пространство V_n наблюдений $x = (x_1, \ldots, x_n)$ и будем считать, что вектор $E(x) = P\Theta$ при гипотезе H_0 лежит в подпространстве V_m, порождаемом вектор-столбцами матрицы P ($E(x) = \eta_{(m)}$); при альтернативных гипотезах $H_\nu, \nu = \overline{1, C_n^k}$, вектор $E(x)$ может принадлежать некоторому из C_n^k возможных подпространств V_{m+k}^ν размерности $(m+k)$. Обозначим через $\hat{x}$ проекцию x на V_m и через $\hat{x}_1$ — проекцию x на $V_{n-m} = V_m^\perp$ (ортогональное дополнение к V_m). Тогда $x = \hat{x} + \hat{x}_1$, поскольку каждый вектор $v \in V_n$ допускает единственное представление вида $v = v_1 + v_2$, где $v_1 \in V_1$, $v_2 \in V_1^\perp$. Вектор $\hat{x}$ является наилучшим приближением вектора x в подпространстве V_m в том смысле, что на нем достигается минимум функции $\|x - x\|^2$ [4]. (Здесь V_n разлагается в прямую сумму подпространств V_m и V_{n-m}: $V_n = V_m \oplus V_{n-m}$).

Рассмотрим структуру подпространств V_{n-m}^ν. Согласно предположениям о сдвиге некоторых наблюдений $x_{\nu_1}, x_{\nu_2}, \ldots, x_{\nu_k}$ относительно выбранной модели, будем идентифицировать с каждой гипотезой H_ν векторное подпространство $V_{k\nu} \subset V_{n-m}^\nu$ и полагать, что вектор $E(\hat{x}^\perp) = \eta_{(m)}^\perp$ лежит в этом подпространстве. Обозначим $\hat{x}_{V_{k\nu}^\perp}$ — проекцию вектора $\hat{x}^\perp$ на $V_{k\nu}$ — через $V_{k\nu}^\perp$ — ортогональное дополнение $V_{k\nu}$ до V_{n-m}^ν. По критерию факторизации статистика $U(\hat{x}, \|\hat{x}^\perp\|^2)$ достаточна для $\mathcal{N}(\eta_{(m)}, \sigma^2 I_n)$, так как при H_0

$$\frac{1}{2\sigma^2}\|x - \eta_{(m)}\|^2 = \frac{1}{2\sigma^2}(\|\hat{x} - \eta_{(m)}\|^2 + \|\hat{x}^\perp\|^2).$$

В то же время при H_ν достаточной статистикой для $\eta_{(m)}^\perp$ будет $T_\nu(x) =$

$= \hat{x}^{\perp}_{V_{kv}}$. Нетрудно заметить, что функции

$$\psi(U, T_v) = \psi_v(x) = \frac{\left\| \hat{x}^{\perp}_{V_{kv}} \right\|^2}{\left\| \hat{x}^{\perp}_{V^{\perp}_{kv}} \right\|^2}$$

монотонно возрастают по T_v при фиксированной U и независимы от U при H_0. Симметричным критерием проверки линейной гипотезы H_0 при $H_v: E(\hat{x}^{\perp}) \in V_{kv}$ будет критерий с критической областью S_k, определяемой неравенством

$$\max_v \psi_v(x) = \max_v \frac{\left\| \hat{x}^{\perp}_{V_{kv}} \right\|^2}{\left\| \hat{x}^{\perp}_{V^{\perp}_{kv}} \right\|^2} \geqslant F_\alpha(k, n-m-k),$$

где $F_\alpha(k, n-m-k)$ — 100α-процентная точка распределения переменной $\max_v \psi_v(x)$. Здесь критическая область S_k является подобной размера α, поскольку функции $\psi_v(x)$ независимы от U и $P_{\theta,\sigma^2}(S_k \mid U = U_0) = \alpha$ для всех $(\theta_1, \theta_2, \ldots, \theta_m)$ и σ^2 (вероятность отклонения гипотезы H_0 равна α на каждой из поверхностей $U = U_0$) [5]. Таким образом, РНМ симметричным критерием в классе подобных является критерий $\varphi(x)$ вида

$$\begin{aligned} &\hat{\varphi}_0(x) = 1, \text{ если } \max_v \psi_v(x) < F_\alpha; \\ &\hat{\varphi}_j(x) = 1, \text{ если } \psi_j(x) = \max_v \psi_v(x) \geqslant F_\alpha. \end{aligned} \tag{8}$$

6. Конструкция прямой суммы подпространств, примененная выше, является теоретической основой дисперсионного анализа, позволяющей эффективно осуществить ортогональное разложение гипотез [6]. Для более полной редукции проблемы отбора выпадающих наблюдений в рассматриваемой потановке введем последовательность «группированных» гипотез: $H_1^* = H^{i_1}$ [в выборке $(x_1, \ldots, x_n)$ имеется один выпадающий элемент — x_{i_1}], $H_2^* = H^{i_1} \cap H^{i_2}$ (два элемента — x_{i_1} и x_{i_2}), ...; $H_k^* = H^{i_1} \cap H^{i_2} \cap \ldots \cap H^{i_k}$. В соответствии с гипотезами $H^{il}, l = \overline{1, k}$, введем k подпространств $W_{i_1}, W_{i_2}, \ldots, W_{i_k}$, попарно ортогональных и порождающих все V_{kv}: $V_{kv} = \bigoplus_{j=1}^{k} W_{vj}$. Очевидно, что $V_n = V_m \oplus V_{n-m}$ и операторами проектирования на V_m и V_{n-m} являются соответственно $P(P^TP)^{-1}P^T$ и $(I - P(P^TP)^{-1}P^T)$. Пусть $B_0 = V^v_{n-m}$ и $B_r = \left(\bigoplus_{i=1}^{r} W_{vi} \right)^{\perp}$, так что $B_k = V^{\perp}_{kv}$ и $B_{k+1} = \varnothing$. Подпространства B_i образуют убывающую последовательность подпространств $V^v_{n-m} = B_0 \supset B_1 \supset \ldots \supset B_k \supset B_{k+1} = \varnothing$. Обозначая через E_i проектор на подпространство B_i, получаем «убывающую» последовательность проекционных операторов

$$I - P(P^TP)^{-1}P^T = E_0 \supset E_1 \supset \ldots \supset E_k \supset E_{k+1} = 0.$$

Поскольку

$$V^v_{n-m} = \bigoplus_{i=0}^{k} (B_i \ominus B_{i+1}) = \left(\bigoplus_{i=1}^{k} W_{vi} \right) \oplus V^{\perp}_{kv}, \tag{9}$$

а оператор ортогонального проектирования на подпространство $(B_i \ominus \ominus B_{i+1})$ равен разности $(E_i - E_{i+1})$, то разложение (9) можно записать

в терминах проекционных операторов:

$$I - P(P^TP)^{-1}P^T = \sum_{i=0}^{k}(E_i - E_{i+1}).$$

Это дает разложение вектора X на сумму $(k+2)$ взаимно ортогональных векторов:

$$x = P(P^TP)^{-1}P^Tx + (I - P(P^TP)^{-1}P^T)x = P(P^TP)^{-1}P^Tx + \\ + \sum_{i=0}^{k}(E_i - E_{i+1})x = \hat{x}_{V_m} + \hat{x}_{V^{\perp}_{k\nu}} + \sum_{j=1}^{k}\hat{x}_{W_{\nu_j}},$$

откуда

$$\|x\|^2 = \|\hat{x}_{V_m}\|^2 + \|\hat{x}_{V^{\perp}_{k\nu}}\|^2 + \sum_{j=1}^{k}\|\hat{x}_{W_{\nu_j}}\|^2.$$

Нетрудно показать, что проекционные операторы равны

$$E_r = \Lambda^{\nu}_r P(P^T\Lambda^{\nu}_r P)^{-1}P^T\Lambda^{\nu}_r,$$

где

$$\Lambda^{\nu}_1 = \begin{pmatrix} & \nu_1 & \\ 1 & \ddots & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix}\begin{matrix}\\ \\ \nu_1 \\ \\ \\ \end{matrix}; \quad \Lambda^{\nu}_2 = \begin{pmatrix} & \nu_1 & \nu_2 & \\ 1 & \ddots & \vdots & 0 \\ & \ddots & 0 & \\ & & 1 & \\ 0 & & \ddots & \\ & & & 1 \end{pmatrix}\begin{matrix}\\ \\ \nu_1 \\ \\ \nu_2 \\ \\ \end{matrix} \quad \text{и т. д.}$$

Используя это обстоятельство, получаем разложение квадратичной формы

$$\|x\|^2 = x^Tx = x^T(P(P^TP)^{-1}P^T)x + x^T(I - P(P^TP)^{-1}P^T - \Lambda^{\nu}_1 + \\ + \Lambda^{\nu}_1 P(P^T\Lambda^{\nu}_1 P)^{-1}P^T\Lambda^{\nu}_1)x + x^T(\Lambda^{\nu}_1 - \Lambda^{\nu}_1 P(P^T\Lambda^{\nu}_1 P)^{-1}P^T\Lambda^{\nu}_1 - \Lambda^{\nu}_2 + \\ + \Lambda^{\nu}_2 P(P^T\Lambda^{\nu}_2 P)^{-1}P^T\Lambda^{\nu}_2)x + \ldots + x^T(\Lambda^{\nu}_{k+1} - \Lambda^{\nu}_{k+1}P(P^T\Lambda^{\nu}_{k+1}P)^{-1} \times \\ \times P^T\Lambda^{\nu}_{k+1})x = \|\hat{x}_{V_m}\|^2 + \|\hat{x}_{V^{\perp}_{k\nu}}\|^2 + Z^2_{\nu_1} + Z^2_{\nu_2} + \ldots + Z^2_{\nu_k} = \\ = Z^2_m + Z^2_{n-m-k} + \sum_{i=1}^{k} Z^2_{\nu_i}, \qquad (10)$$

на котором основывается применение следующей процедуры отбора выпадающих наблюдений.

7. Будем считать, что независимые переменные $Z_{(i_1)} < Z_{(i_2)} < \ldots < Z_{(i_k)}$ выбраны таким образом, что они соответствуют k «подозреваемым» наблюдениям $x_{i_1}, \ldots, x_{i_k}$, имеющим максимальные отклонения от их оцененных значений (это можно сделать последовательно). Нетрудно заметить, что $E(Z_{(i_j)}) = \eta^{\perp}_j$ и $\max_{\nu}\psi_{\nu}(x) = \sum_{j=1}^{k} Z^2_{(i_j)}/Z^2_{n-m-k}$.

Если число выпадающих наблюдений неизвестно, но не больше k, то нужно решить, к какому из k подпространств $V_{r\nu} = \bigoplus_{j=1}^{r} W_{\nu_j}$ принадлежит вектор $E(Z^{\nu}_k) = (E(Z_{\nu_1}), E(Z_{\nu_2}), \ldots, E(Z_{\nu_k}))$. Для поиска соответствую-

щей процедуры формализуем эту задачу известным образом [7], считая

$$
\begin{aligned}
&H_k^* = \hat{H}^{i_1} : \eta_1^{\perp} \neq 0 \quad (\eta^{\perp} \in V_{ki});\\
&H_{k-1}^* = \hat{H}^{i_2} : \eta_1^{\perp} = 0, \quad \eta_2^{\perp} \neq 0 \quad (\eta^{\perp} \in V_{(k-1)i});\\
&\vdots\\
&H_1^* = \hat{H}^{i_k} : \eta_1^{\perp} = \eta_2^{\perp} = \ldots = \eta_{k-1}^{\perp} = 0, \quad \eta_k^{\perp} \neq 0 \quad (\eta^{\perp} \in W_{i_1});\\
&\hat{H}^{i_{k+1}} : \eta_1^{\perp} = \eta_2^{\perp} = \ldots = \eta_k^{\perp} = 0.
\end{aligned}
\tag{11}
$$

Данной последовательности гипотез противопоставим последовательность нулевых гипотез:

$$
\begin{aligned}
&\hat{H}_0^{i_1} : \eta_1^{\perp} = 0;\\
&\hat{H}_0^{i_1 i_2} : \eta_1^{\perp} = \eta_2^{\perp} = 0;\\
&\vdots\\
&\hat{H}_0^{i_1 i_2 \ldots i_k} : \eta_1^{\perp} = \eta_2^{\perp} = \ldots = \eta_k^{\perp} = 0.
\end{aligned}
\tag{12}
$$

Ясно, что если некоторая гипотеза $H_0^{i_1 \ldots i_l}$ верна, то предшествующие ей гипотезы также должны быть верны. Если ж она неверна, то неверны и все последующие гипотезы: $(\hat{H}_0^{i_1} \supset \hat{H}_0^{i_1 i_2} \supset \ldots \supset \hat{H}_0^{i_1 i_2 \ldots i_k})$. Кроме того, семейства множеств, определяемые гипотезами (11) и (12), связаны соотношениями:

$$
\begin{aligned}
&\hat{H}_0^{i_1} = \hat{H}^{i_2} : \eta^{\perp} \in W_{i_2} \oplus W_{i_3} \oplus \ldots \oplus W_{i_k};\\
&\hat{H}_0^{i_1 i_2} = \hat{H}^{i_3} : \eta^{\perp} \in W_{i_3} \oplus W_{i_4} \oplus \ldots \oplus W_{i_k};\\
&\vdots\\
&\hat{H}_0^{i_1 i_2 \ldots i_k} = \hat{H}^{i_{k+1}} : \eta^{\perp} \in \varnothing.
\end{aligned}
\tag{13}
$$

Здесь будем ограничивать вероятности ошибок, связанных с принятием решений о том, что коэффициенты $\eta_i^{\perp}, i = \overline{1, k}$, отличны от нуля, в то время как на самом деле они равны нулю. С этой целью зададим каждой из нулевых гипотез определенный уровень значимости:

$$
\begin{aligned}
&\pi_1 = P\{\text{отвергнуть } \hat{H}_0^{i_1} \mid \hat{H}_0^{i_1}\};\\
&\pi_1 + \pi_2 = P\{\text{отвергнуть } \hat{H}_0^{i_1 i_2} \mid H_0^{i_1 i_2}\};\\
&\vdots\\
&\pi_1 + \pi_2 + \ldots + \pi_k = P\{\text{отвергнуть } \hat{H}_0^{i_1 i_2 \ldots i_k} \mid \hat{H}_0^{i_1 i_2 \ldots i_k}\} =\\
&\qquad = P\{\text{отвергнуть } \hat{H}^{i_{k+1}} \mid \hat{H}^{i_{k+1}}\},
\end{aligned}
\tag{14}
$$

где $\pi_i \geqslant 0$ и $\sum_{i=1}^{k} \pi_i \leqslant 1$. В такой постановке вероятность отклонения каждой последующей нулевой гипотезы, когда она верна, не меньше, чем аналогичная вероятность для предшествующей гипотезы. Использование этих соотношений дает определенные значения вероятностям принятия гипотезы $\hat{H}^{i_l}$, когда на самом деле верна одна из последующих гипотез $\hat{H}^{i_{l+1}}, \ldots, \hat{H}^{i_k}$:

$$
\begin{aligned}
&\pi_1 = P\{\text{принять } \hat{H}^{i_1} \mid \hat{H}^{i_2} \cup \ldots \cup \hat{H}^{i_{k+1}}\};\\
&\pi_2 = P\{\text{принять } \hat{H}^{i_2} \mid \hat{H}^{i_3} \cup \ldots \cup \hat{H}^{i_{k+1}}\};\\
&\vdots\\
&\pi_k = P\{\text{принять } \hat{H}^{i_k} \mid \hat{H}^{i_{k+1}}\}.
\end{aligned}
\tag{15}
$$

Как было показано выше, критерий (8) с областью отклонения S_k является оптимальным для проверки гипотезы $\hat{H}_0^{i_1} : \eta_1^{\perp} = 0$ (дополнительной

$\ldots \widehat{H}^{i_1} - H^*_{(i_l)}$ . ... когда число b известно. Если же b не-

для проверки гипотезы $\widehat{H}_0^{i_1 \ldots i_l}: \eta_l^{\perp} = 0$ — с уровнем значимости α в предположении, что $\eta_1^{\perp} = \ldots = \eta_{l-1}^{\perp} = 0$, состоит в ее отклонении, если

$$t_l^2 = \frac{Z^2_{(i_l)}}{Z^2_{n-m-k} + Z^2_{(i_1)} + \ldots + Z^2_{(i_{l-1})}} \geqslant \frac{t_l^2(\alpha)}{n-m-k+l+1}. \qquad (16)$$

Здесь вероятность отклонения гипотезы $\widehat{H}_0^{i_1, i_2 \ldots i_l}$ не зависит от «мешающих» параметров $\theta_1, \theta_2, \ldots, \theta_m, \sigma^2$ и $\eta_{l+1}^{\perp}, \ldots, \eta_k^{\perp}$. Тем самым, если S_l — подобная область размера α_l для проверки гипотезы $\widehat{H}_0^{i_1 i_2 \ldots i_l}: \eta_l^{\perp} = 0$, или формально

$$P\{s_l \mid \eta_1^{\perp} = \eta_2^{\perp} = \ldots = \eta_{l-1}^{\perp} = \eta_l^{\perp} = 0\} = \alpha_l$$

для всех $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_m)^T$, σ^2, то S_l выделяет условную вероятность α_l на почти всех комбинациях значений достаточных для этих параметров статистик $\widehat{x}_{V_m}$ и $z^2_{n-m-k+l}$ при $\eta_1^{\perp} = \eta_2^{\perp} = \ldots = \eta_l^{\perp} = 0$.

Нерандомизированная статистическая процедура для данной задачи состоит в следующем [7]. Фиксируется набор $(k+1)$ попарно непересекающихся областей $R_1, R_2, \ldots, R_k, R_{k+1}$ в выборочном пространстве V_n. Если выборочная точка попадает в R_j, то принимается гипотеза $\widehat{H}^{ij} = H^*_{k-j+1}$. Приписывание уровней значимости (15) приводит к тому, что эти области становятся «подобными» в том смысле, что при $\eta_1^{\perp} = \eta_2^{\perp} = \ldots = \eta_l^{\perp} = 0$ вероятности попадания выборочной точки в $R_1, R_2, \ldots, R_l$ равны соответственно $\pi_1, \pi_2, \ldots, \pi_l$ независимо от $(\Theta_1, \Theta_2, \ldots, \Theta_m)$, σ^2 и $\eta_{l+1}^{\perp}, \eta_{l+2}^{\perp}, \ldots, \eta_k^{\perp}$. Пусть T_l^* — область, определяемая соотношением (16) при $\alpha_l = \pi_l/(1-\pi_1-\pi_2-\ldots-\pi_{l-1})$. Тогда согласно [7], какими бы ни были $R_1, R_2, \ldots, R_{l-1}$, наилучший выбор области R_l состоит в том, что она должна являться частью множества T_l^*, не содержащей точек из $R_1, R_2, \ldots, R_{l-1}$. При таком выборе $(R_l = W_{il})$ вероятность

$$P\{R_l \mid \eta_1^{\perp} = \eta_2^{\perp} = \ldots = \eta_{l-1}^{\perp} = 0\} = (1 - \pi_1 - \pi_2 - \ldots - \pi_{l-1})\, P\{T_l^* \mid \eta_1^{\perp} = \ldots = \eta_l^{\perp} = 0\}$$

не зависит от выбора $R_1, \ldots, R_{l-1}$ и для каждого значения $\eta_l^{\perp}$ вероятность $P\{R_l \mid \eta_l^{\perp}, \eta_1^{\perp} = \ldots = \eta_{l-1}^{\perp} = 0\}$ принимает максимальное значение. Оптимальная процедура состоит, по сути дела, в поочередной проверке гипотез $\eta_1^{\perp} = 0$, $\eta_2^{\perp} = 0, \ldots$ до тех пор, пока либо какая-то из гипотез не будет отвергнута, либо будут приняты все гипотезы вплоть до $\widehat{H}_0^{i_1 i_2 \ldots i_k} = \widehat{H}^{i_k+1}$. Если все α_l фиксированы и равны α, то $\pi_1 = \alpha$, $\pi_2 = \alpha(1-\alpha)$, $\ldots$, $\pi_k = \alpha(1-\alpha)^k$, $\pi_{k+1} = (1-\alpha)^{k+1}$. Наконец, удобные для практических расчетов аппроксимации распределения статистик t_l^2 можно получить, используя результаты статьи [8].

ЛИТЕРАТУРА

1. Ferguson T. S. On the rejection of outliers.— Proc. of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. California, University of California, Press, 1961, vol. 1, p. 253—287.
2. Tietjen G. L., More R. H. Some Grubbs-type statistics for the detection of several outliers.— "Technometrics", 1972, vol. 14, N 3, p. 583—597.
3. Киричук В. С., Луценко Б. Н. Исключение недостоверных данных.— «Автометрия», 1970, № 6, с. 47—54.
4. Рао С. Р. Линейные статистические методы и их применения. М., «Наука», 1968.
5. Леман Е. Л. Проверка статистических гипотез. М., «Наука», 1964.
6. Барра Ж.-Р. Основные понятия математической статистики. М., «Мир», 1974.
7. Андерсон Т. Статистический анализ временных рядов. М., «Мир», 1976.
8. Большев Л. Н. О критериях исключения резко выделяющихся наблюдений.— Труды Ин-та прикладной математики Тбилисского гос. ун-та. Вып. II. Тбилиси, 1969, с. 159—177.

Поступила в редакцию 21 января 1976 г.