Метод формирования выборок непараметрических алгоритмов для обеспечения независимости распределений их решающих статистик при гипотезе

Р. В. Подрезов*, М. А. Райфельд

Новосибирский государственный технический университет, Россия

*Контактный e-mail: podrezov-r.v@mail.ru

Технологический прорыв в создании MEMS-устройств — миниатюрных датчиков движения интегрального исполнения (акселерометров, гироскопов, магнитных компасов) привел к развитию методов инерциальной навигации, основанных на обработке сигналов этих датчиков. В статье предложено решение задачи надежного различения состояний шахтного лифта движение/останов, основанное на обработке сигналов MEMS-акселерометра, установленного в лифтовой кабине. Обработка сигнала акселерометра представляет собой комбинацию двух непараметрических алгоритмов: критерия, основанного на ранговых статистиках Вилкоксона, и критерия превышающих наблюдений. Приведено доказательство возможности такого формирования выборок и ранжирования, при которых решающие статистики этих критериев независимы.

Ключевые слова: непараметрический критерий, формирование выборки, цифровая обработка сигналов.

Введение

В статье рассматривается подход к построению непараметрического критерия, точнее, совместного использования нескольких непараметрических критериев (или совместного критерия) для принятия решений в случае нестационарного сигнала, распределение которого при альтернативе на разных временных интервалах отличается от распределения при гипотезе такими параметрами, как сдвиг, масштаб и другие.

Данная задача является актуальной в системах инерциальной навигации, основанных на использовании датчиков ускорения (акселерометров). Поскольку объект в разные интервалы времени может находиться в состоянии покоя, ускоренного или инерционного движения и эти интервалы чередуются с определенной закономерностью, соответствующими свойствами обладают и сигналы, регистрируемые датчиками, которые в этом случае являются нестационарными.

Примером подобного сигнала может служить сигнал, регистрируемый акселерометром, установленным в лифтовой кабине, и используемый для принятия решения о начале ее движения либо об остановке [1]. Временна́я реализация данного сигнала $s(t)$ при гипотезе H_0 (кабина лифта неподвижна) представляет собой шум акселерометра — стационарный процесс, характеризующийся нулевым математическим ожиданием ($m_0 = 0$)

и некоторой дисперсией σ_0^2. Распределение этого процесса в общем случае не является гауссовским. При альтернативе (начало движения кабины) можно выделить несколько временны́х интервалов, на каждом из которых сигнал акселерометра характеризуется своим распределением (рис. 1).

Наличие этих временны́х интервалов в сигнале акселерометра обусловлено физическими процессами, имеющими место при разгоне либо торможении лифтовой кабины. Так, в случае разгона интервал *1* связан с отсутствием регулярного силового воздействия на кабину (лифт стоит на месте, ускорение равно нулю), интервал *2* — период ускорения кабины за счет внешней силы (в сигнале акселерометра появляется смещение, которое говорит о наличии постоянного ускорения), интервал *3* — лифт движется по инерции, ускорение равно нулю (более высокая дисперсия сигнала связана с вибрацией при движении кабины). При торможении кабины указанные временны́е интервалы следуют в обратном порядке, а импульс ускорения имеет другую полярность.

Из рис. 1 следует, что при альтернативе выборка наблюдений, извлеченная из области *2*, отличается сдвигом Δ от выборок в областях *1* и *3*, а выборка, извлеченная из области *3*, отличается масштабным коэффициентом μ от наблюдений в области *1*. Если считать вид распределения наблюдений неизвестным, но единым для всех выборок, а отличия связаны только с параметрами этого распределения, то формальное определение связи распределений при альтернативе и гипотезе можно представить в следующем виде:

$$\begin{aligned} F_{01}(x) &= F_{11}(x), \\ F_{02}(x) &= F_{12}(x-\Delta), \\ F_{03}(x) &= F_{13}(\mu x). \end{aligned} \tag{1}$$

В последнем выражении первый индекс в обозначении функции распределения указывает на то, что оно соответствует гипотезе (0) или альтернативе (1), второй индекс — на номер временной области сигнала в соответствии с рис. 1. При гипотезе H_0 все три выборки подчиняются единому распределению $F_0(x)$. Поскольку вид этого распределения неизвестен, а при принятии решения требуется обеспечивать стабильные рабочие характеристики (по критерию Неймана—Пирсона), в данном случае оправдано использование двухвыборочных непараметрических критериев, основанных на ранговых ста-

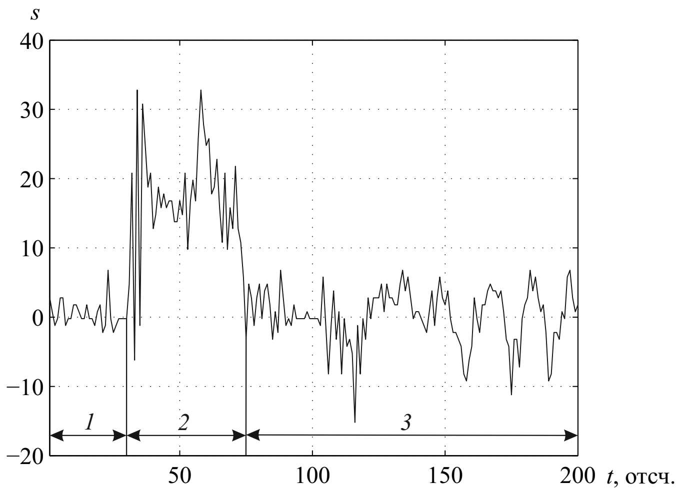

Рис. 1

тистиках. Обозначим $X = \{x_1, x_2, \ldots, x_n\}$ общую выборку, включающую подвыборки $X_1 = \{x_{11}, x_{12}, \ldots, x_{1m}\}$, $X_2 = \{x_{21}, x_{22}, \ldots, x_{2k}\}$ и $X_3 = \{x_{31}, x_{32}, \ldots, x_{3l}\}$. Объем общей выборки составляет $n = m + k + l$. Поскольку общая выборка при альтернативе не является стационарной и включает области, отличающиеся сдвигом и масштабным коэффициентом, алгоритм принятия решения о начале движения кабины лифта можно построить с использованием статистик двух непараметрических тестов, ориентированных на обнаружение соответствующих признаков, характеризующих участки общей выборки X.

В качестве рангового критерия, "чувствительного" к наличию сдвига в распределениях, можно использовать двухвыборочный тест Вилкоксона [2] с решающей статистикой S:

$$S = \sum_{i=1}^{m} R_i, \tag{2}$$

где R_i, $i = \overline{1, m}$, — ранги элементов рабочей выборки $Y = \{y_1, y_2, \ldots, y_m\}$, относительно наблюдений которой и проверяется гипотеза о наличии сдвига. Ранжирование при этом осуществляется по общей выборке X объема n элементов, частью которой является и рабочая выборка. Другая часть общей выборки $Z = \{z_1, z_2, \ldots, z_{n-m}\}$ объема $n-m$ элементов называется опорной выборкой. Она содержит отсчеты, распределение которых соответствует гипотезе нулевого сдвига.

В качестве критерия, ориентированного на обнаружение масштабных различий выборок, можно использовать двухвыборочный критерий превышающих наблюдений [3] с решающей статистикой T:

$$T = \sum_{i=1}^{m} \delta_i,$$

$$\delta_i = \begin{cases} 0, & z_{\min} \le y_i \le z_{\max}, \\ 1, & y_i < z_{\min} \ \text{ или } \ y_i > z_{\max}. \end{cases} \tag{3}$$

Статистику T можно определить и на основе ранговых статистик опорной выборки: $T = n - (R_{\max} - R_{\min} + 1)$. Здесь $R_{\max}$ и $R_{\min}$ — максимальный и минимальный ранги элементов опорной выборки Z. Решение принимается на основе анализа непараметрических статистик S и T. В общем случае эти статистики не являются независимыми, поскольку вычисляются на основе одних и тех же элементов общей выборки X. Таким образом, их совместное распределение $P(S, T)$ не представляется в виде произведения одномерных плотностей $P(S)$ и $P(T)$, что значительно усложняет построение алгоритма принятия решения. Однако можно показать, что, следуя определенному правилу при построении выборок критериев, можно обеспечить независимость статистик S и T (и любых ранговых статистик) при гипотезе H_0 в случае их вычисления по единой выборке наблюдений.

Описание критерия

Пусть имеется общая выборка X, состоящая из n элементов. Эта выборка разбита на несколько подвыборок: X_1 размера n_1, X_2 размера n_2, ..., X_k размера n_k (рис. 2). При этом $n = n_1 + n_2 + \ldots + n_k$. Можно доказать, что распределение рангов рабочей выборки (а следовательно, и любых статистик, построенных на основе этих рангов) из X_2,

X_1, n_1	X_2, n_2	X_3, n_3	- - - - - - - - - - - -	X_k, n_k

Рис. 2

полученных для общей выборки, составленной из элементов X_1 и X_2, не зависит от распределения рангов рабочей выборки X_3, полученных для общей выборки, составленной из элементов X_1, X_2 и X_3, и эти распределения не зависят от распределения рангов рабочей выборки X_k, полученных для общей выборки, составленной из элементов $X_1, X_2, X_3, \ldots, X_k$. Иначе говоря,

$$P(R_2, R_3, \ldots, R_k|H_0) = P_1(R_2|H_0) \cdot P_2(R_3|H_0) \cdot \ldots \cdot P_{k-1}(R_k|H_0), \tag{4}$$

где $R_2 = \{r_{21}, r_{22}, \ldots, r_{2n_2}\}$ — ранги элементов выборки $X_2 = \{x_{21}, x_{22}, \ldots, x_{2n_2}\}$, полученных для общей выборки $\{X_1, X_2\}$, $R_3 = \{r_{31}, r_{32}, \ldots, r_{3n_3}\}$ — ранги элементов выборки $X_3 = \{x_{31}, x_{32}, \ldots, x_{3n_3}\}$, полученных для общей выборки $\{X_1, X_2, X_3\}$, и $R_k = \{r_{k1}, r_{k2}, \ldots, r_{kn_k}\}$ — ранги элементов выборки $X_k = \{x_{k1}, x_{k2}, \ldots, x_{kn_k}\}$, полученных для общей выборки $\{X_1, X_2, X_3, \ldots, X_k\}$.

Для доказательства приведенного выше утверждения покажем, что распределение рангов рабочей выборки $\mathbf{R}_Y = \{r_{y1}, r_{y2}, \ldots, r_{yn_2}\}$, $Y = \{y_1, y_2, \ldots, y_{n_2}\} = X_2$, рассчитанных по выборке $X = \{X_1, X_2\}$, не зависит от распределения рангов выборки X_1, рассчитанных по выборке X_1. Пусть вариационный ряд, составленный из элементов общей выборки X, имеет, например, вид

$$X_R = \{z^{(1)}, y^{(2)}, \ldots, y^{(k)}, z^{(k+1)}, \ldots, z^{(i)}, y^{(i+1)}, \ldots, y^{(n_1+n_2-1)}, z^{(n_1+n_2)}\},$$

i-й элемент ряда представляет собой соответствующую (i-ю) порядковую статистику. Верхний индекс указывает номер элемента в общем вариационном ряду, а символ y либо z — принадлежность данного элемента рабочей или опорной выборке, $Z = \{z_1, z_2, \ldots, z_{n_1}\} = X_1$ — элементы опорной выборки. Нижний индекс элемента выборок $X, Y, Z, \mathbf{R}_X, \mathbf{R}_Y, \mathbf{R}_Z$ связан с его временным положением в данной выборке, т. е. моментом его получения. Имеем следующие упорядоченные множества рангов элементов рабочей и опорной выборок $2, \ldots, k, l+1, \ldots, n_1+n_2-1$ и $1, \ldots, k+1, \ldots, l, \ldots, n_1+n_2$. Вероятность рангового вектора общей выборки $\mathbf{R}_X = \{\mathbf{R}_Y, \mathbf{R}_Z\}$, соответствующего вариационному ряду X_R, можно представить в виде

$$P(\mathbf{R}_Y, \mathbf{R}_Z) = \int\limits_{-\infty}^{\infty} P_0(z^{(n_1+n_2)}) dz^{(n_1+n_2)} \int\limits_{-\infty}^{z^{(n_1+n_2)}} P_0(y^{(n_1+n_2-1)}) dy^{(n_1+n_2-1)} \ldots$$
$$\ldots \int\limits_{-\infty}^{y^{(l+1)}} P_0(z^{(l)}) dz^{(l)} \ldots \int\limits_{-\infty}^{z^{(k+1)}} P_0(y^{(k)}) dy^{(k)} \ldots \int\limits_{-\infty}^{y^{(2)}} P_0(z^{(1)}) dz^{(1)}, \tag{5}$$

где $P(\bullet)$ — распределение наблюдений при гипотезе.

Зафиксируем какой-либо ранговый вектор $\mathbf{R}_Y = \{r_{y1}, r_{y2}, \ldots, r_{yn_2}\}$. Тогда всевозможные вариации рангового вектора $\mathbf{R}_X$ связаны с перестановками одной и той же последовательности рангов внутри выборки $\mathbf{R}_Z = \{r_{z1}, r_{z2}, \ldots, r_{zn_1}\}$ (поскольку последовательность $\mathbf{R}_Y$ зафиксирована). Если, например, $n_1 = n_2 = 3$ и $\mathbf{R}_Y = \{3, 1, 5\}$, то возможными значениями, принимаемыми рангами опорной выборки (при ранжировании в общей выборке $X = \{Z, Y\}$ размера $n_1 + n_2 = 6$), являются $\{2, 4, 6\}$, $\{2, 6, 4\}$, $\{4, 2, 6\}$, $\{6, 2, 4\}$, $\{4, 6, 2\}$, $\{6, 4, 2\}$. Эти значения соответствуют следующим ранговым последовательностям, построенным по наблюдениям опорной выборки Z: $\mathbf{R}_{Z1} = \{1, 2, 3\}, \mathbf{R}_{Z2} = \{1, 3, 2\}, \mathbf{R}_{Z3} = \{2, 1, 3\}$, $\mathbf{R}_{Z4} = \{2, 3, 1\}, \mathbf{R}_{Z5} = \{3, 1, 2\}, \mathbf{R}_{Z6} = \{3, 2, 1\}$ (т. е. представляющим собой все $m = n!$ перестановок рангов выборки $\mathbf{R}_Z$). Очевидно, что перестановка элементов в выборке $\mathbf{R}_Z$ не влияет на значения порядковых статистик, а следовательно, на результаты расчета $P(\mathbf{R}_Y, \mathbf{R}_Z)$ по формуле (5). Поэтому $P(\mathbf{R}_Y, \mathbf{R}_{Z1}) = P(\mathbf{R}_Y, \mathbf{R}_{Z2}) = \ldots = P(\mathbf{R}_Y, \mathbf{R}_{Z6})$. Если теперь учесть, что распределение $P(\mathbf{R}_Z)$ также не зависит от перестановок элементов выборки $\mathbf{R}_Z$ и $P(\mathbf{R}_Z) = 1/n_1!$, то с учетом того, что $P(\mathbf{R}_Y, \mathbf{R}_Z)/P(\mathbf{R}_Z) = P(\mathbf{R}_Y|\mathbf{R}_Z)$, имеем

$$P(\mathbf{R}_Y|\mathbf{R}_{Z1}) = P(\mathbf{R}_Y|\mathbf{R}_{Z2}) = \ldots = P(\mathbf{R}_Y|\mathbf{R}_{Zm}). \tag{6}$$

Все возможные перестановки выборки Z составляют полную группу событий.

Таким образом, вероятность заданного рангового вектора $\mathbf{R}_Y$ — одна и та же для любой перестановки рангового вектора $\mathbf{R}_Z$, поэтому она не зависит от $\mathbf{R}_Z$ и, следовательно, $P(\mathbf{R}_Y|\mathbf{R}_Z) = P(\mathbf{R}_Y)$. Далее в качестве рабочей примем выборку X_3 (рис. 2), $Y = X_3$, а в качестве опорной используем объединение $Z = \{Z_1, Z_2\}$, где $Z_1 = X_1, Z_2 = X_2$. Имеем

$$P(\mathbf{R}_Y, \mathbf{R}_Z) = P(\mathbf{R}_Y, \mathbf{R}_{Z1}, \mathbf{R}_{Z2}) = P(\mathbf{R}_Y|\mathbf{R}_Z)P(\mathbf{R}_{Z2}|\mathbf{R}_{Z1})P(\mathbf{R}_{Z1}). \tag{7}$$

С учетом (6)

$$P(\mathbf{R}_Y, \mathbf{R}_{Z1}, \mathbf{R}_{Z2}) = P(R_Y)P(\mathbf{R}_{Z2})P(\mathbf{R}_{Z1}), \tag{8}$$

что говорит о независимости ранговых последовательностей соответствующих выборок.

Вернемся к задаче об обнаружении начала движения лифтовой кабины. Сформируем выборки непараметрических алгоритмов (2), (3) в соответствии с изложенным выше правилом, целью которого является обеспечение независимости статистик S и T. Выборки, изображенные на рис. 3, соответствуют временны́м интервалам сигнала акселерометра, имеющего место при старте лифтовой кабины (см. рис. 1). В качестве опорной выборки используем X_1, которая содержит только шум акселерометра, находящегося в покое. Выборка X_2, соответствующая интервалу ускорения, распределение которой характеризуется сдвигом (при альтернативе), рассматривается как рабочая выборка алгоритма Вилкоксона. Опорная выборка этого алгоритма включает области $\{X_1, X_3\}$. Выборка X_3, состоящая из отсчетов сигнала инерционного движения лифта, является рабочей выборкой алгоритма превышающих наблюдений, в качестве опорной выборки этого алгоритма используется X_1. При гипотезе все эти выборки подчиняются одному распределению и поэтому

$$P(T, S|H_0) = P(S|H_0)P(T|H_0). \tag{9}$$

Если теперь рассмотреть статистики T и S, вычисляемые для общей опорной выборки X_1, то они окажутся зависимыми. Можно показать, что $P(S|T, H_0) \neq P(S|H_0)$.

X_1, n_1	X_2, n_2	X_3, n_3

Рис. 3

Для этого возьмем простейший случай, когда выборка содержит всего два отсчета (чтобы была возможность вычислить $x_{\min}$ и $x_{\max}$ в соответствии с выражением (3)), а X_2 и X_3 — по одному отсчету. В этом случае распределение статистики Вилкоксона является равномерным $P(S|H_0) = 1/3$ в интервале целых чисел $[1, 3]$, поскольку S фактически представляет собой ранг элемента X_2, вычисляемый для выборки $\{X_1, X_2\}$. Предположим, что значение статистики T, полученное для той же опорной выборки X_1, равно 1 ($T = 1$). В этом случае $P(S = 3|T = 1, H_0)$ вычисляется следующим образом:

$$P(S = 3, T = 1) = 2\int\limits_{-\infty}^{\infty} P_0(x_{11})dx_{11}\int\limits_{x_{11}}^{\infty} P_0(x_{12})dx_{12}\int\limits_{x_{12}}^{\infty} P_0(x_{31})dx_{31}\int\limits_{x_{12}}^{\infty} P_0(x_{21})dx_{21}+$$
$$+2\int\limits_{-\infty}^{\infty} P_0(x_{11})dx_{11}\int\limits_{x_{11}}^{\infty} P_0(x_{12})dx_{12}\int\limits_{-\infty}^{x_{11}} P_0(x_{31})dx_{31}\int\limits_{x_{12}}^{\infty} P_0(x_{21})dx_{21} = 1/4. \qquad (10)$$

Поскольку вероятность $P(T = 1|H_0) = 2/3$, получаем $P(S = 3|T = 1, H_0) = (1/4)/(2/3) = 3/8$, что не совпадает с $P(S|H_0)$. В результате расчетов, выполненных по формуле (10), получены данные, которые говорят о статистической зависимости T и S при использовании общей опорной выборки X_1:

S	1	2	3
$P(S\|H_0)$	1/3	1/3	1/3
$P(S\|T = 1, H_0)$	3/8	1/4	3/8
$P(S\|T = 0, H_0)$	1/4	1/2	1/4

В то время как распределение статистики превышающих наблюдений при гипотезе задается явным выражением [2]

$$P(T|H_0) = \frac{C_m^T n(n-1)}{C_{m+n}^{T+2}(T+2)}, \qquad (11)$$

где m и n — соответственно объемы рабочей и опорной выборок, распределение классической статистики Вилкоксона представляется только в виде таблицы из работы [4].

Синтез оптимального совместного алгоритма, использующего статистики S и T, является сложной задачей, получить решение которой в общем виде невозможно (из-за сложной формы распределений). Если аппроксимировать распределения статистик гауссовскими кривыми (такая аппроксимация, конечно, является очень грубой), то оптимальный алгоритм принятия решения сводится к сравнению с порогом C_Q взвешенной суммы Q статистик S и T:

$$Q = a_S S + a_T T. \qquad (12)$$

Порог может выбираться, например, на основе критерия Неймана—Пирсона. Коэффициенты $\{a_S, a_T\}$ зависят, в частности, от дисперсии и математического ожидания распределений соответствующих статистик при гипотезе и альтернативе. Для расчета порога C_Q по критерию Неймана—Пирсона необходимо знать распределение $P(Q|H_0)$, которое рассчитывается по формуле (12), и совместное распределение $P(T, S|H_0)$. В случае, когда совместное распределение S и T представляется в виде произведения (9), $P(Q|H_0)$ можно записать в виде

$$P(Q|H_0) = \int\limits_S P_S(S|H_0) P_T\left(\frac{Q - a_S S}{a_T}|H_0\right) dS. \qquad (13)$$

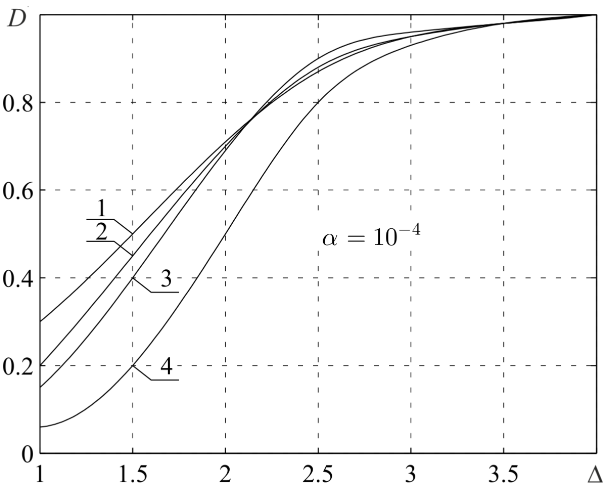

Рис. 4. Зависимость вероятности правильного обнаружения от сдвига: кривые *1–3* — совместные алгоритмы (Вилкоксона и превышающих наблюдений) при $\mu = 3$; 2.5; 2 соответственно; кривая *4* — алгоритм Вилкоксона, $\mu = 2.5$

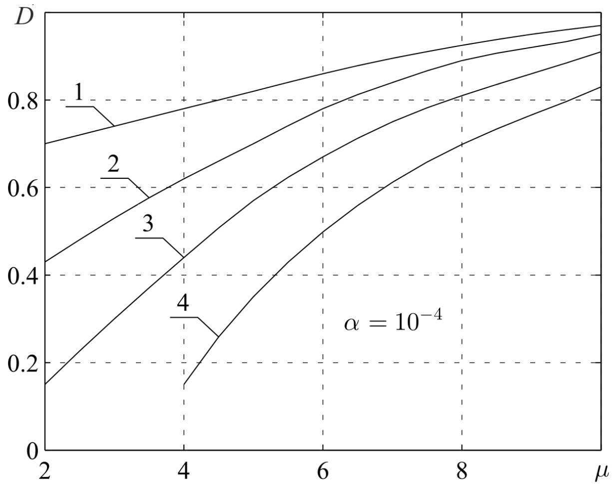

Рис. 5. Зависимость вероятности правильного обнаружения от масштаба: кривые *1–3* — совместные алгоритмы (Вилкоксона и превышающих наблюдений) при $\Delta = 2$; 1.5; 1 соответственно; кривая *4* — алгоритм превышающих наблюдений, $\Delta = 1.5$

Другой подход к построению совместного алгоритма не претендует на оптимальность, но он более прост в реализации. В качестве решающей статистики используется следующая функция:

$$Q = \theta(S - C_S) + \theta(T - C_T), \tag{14}$$

где $\theta(\bullet)$ — функция Хевисайда; C_S, C_T — пороги обнаружения алгоритмов (2) и (3) соответственно. Решение принимается в результате сравнения статистики Q с порогом $C_Q = 2$. При $Q = 2$ принимается решение H_1^*, в противном случае — H_0^*. Фактически решение выносится в результате логического (по И) объединения решений алгоритмов Вилкоксона и превышающих наблюдений. В этом случае достаточно просто рассчитать вероятность ложной тревоги α совместного алгоритма, поскольку она представляется в виде произведения вероятностей ложных тревог отдельных алгоритмов

$$\alpha = \int\limits_{C_T}^{\infty}\int\limits_{C_S}^{\infty} P(T, S|H_0)dTdS = \int\limits_{C_T}^{\infty} P(S|H_0)dS \int\limits_{C_T}^{\infty} P(T|H_0)dT = \alpha_S\alpha_T, \tag{15}$$

а следовательно, и независимо выбрать их пороги обнаружения C_S и C_T. Отметим, что в случае статистической зависимости статистик S и T этого сделать нельзя. В соответствии с указанным подходом к построению совместного критерия (14) рассчитаны рабочие характеристики алгоритма обнаружения старта лифтовой кабины, приведенные на рис. 4 и 5.

Таким образом, предложен и обоснован метод формирования выборок совместных (по нескольким признакам) непараметрических критериев, приводящий к независимости их решающих статистик при гипотезе. На основе данного метода решается задача расчета распределения статистики Вилкоксона при гипотезе. Обосновывается преимущество использования независимых решающих статистик при построении совместного непараметрического критерия.

Благодарности. Работа выполнена при финансовой поддержке Министерства образования и науки РФ по государственному заданию № 2014/138, проект № 1176.

Список литературы / References

[1] **Райфельд М.А.** Различение состояний движение/останов лифта на основе сигналов акселерометра // Автометрия. 2015. № 2(51). С. 93–102.
Rajfeld, M.A. Discrimination between the move and stop of a lift based on accelerometer signals // Avtometriya. 2015. No. 2(51). P. 93–102. (In Russ.)

[2] **Гаек Я., Шидак З.** Теория ранговых критериев / Пер. с англ. М.: Наука, 1971. 376 с.
Hajek, J., Sidak, Z. Theory of rank tests. 2nd edition. N.Y.: Acad. Press, 1999. 435 p.

[3] **Райфельд М.А.** Использование группировки для увеличения мощности непараметрического критерия, основанного на превышающих наблюдениях // Изв. вузов России. Радиоэлектроника. 2006. № 2. С. 28–35.
Rajfeld, M.A. Grouping for the increase of capacity of non parametric criterion based on the elevated observations // Izvestiya Vuzov Rossii. Radio-elektronika. 2006. No. 2. P. 28–35. (In Russ.)

[4] **Большев Л.Н., Смирнов Н.В.** Таблицы математической статистики. М.: Наука. Гл. ред. физ.-мат. лит., 1983. 416 с.
Bol'shev L.N., Smirnov N.V. Tables for mathematical statistics. Moscow: Nauka. Glavnaya redaktsiya fiziko-matematicheskom literatury, 1983. 416 p. (In Russ.)

Поступила в редакцию 5 ноября 2015 г.

A nonparametric test sampling method providing independent distribution of the test statistics under the null-hypothesis

Podrezov, Roman V.*, Rajfeld, Mikhail A.

Novosibirsk State Technical University, Novosibirsk, 630073, Russia

*Corresponding author: Podrezov, Roman V., e-mail: podrezov-r.v@mail.ru

Signal detection problem is usually considered as a testing of statistical hypothesis. When minimal assumptions are made about the distributions, application of a nonparametric test may have considerable advantages in efficiency. Situation becomes more complex in case of non-stationary signal, e. g. signal differs from noise by variance at one time interval and by level at another interval. Thus, the joint hypothesis testing the scale and the shift is required. In practice, nonparametric signal detectors are often developed using a reference sample that represents only noise at some time interval. Two tests have dependent statistics distributions, if they use the same reference sample. In another case, usage of the different reference samples reduces efficiency of a test, because the efficiency is related to minimal sample sizes required to obtain given constraints of the error probabilities.

For rank tests, a developed sampling and ranking method can be applied to make ranks in working samples to be independent, and consequently rank statistics distributions. Ranks of working samples are calculated on general sample that consists of current

sample, previous samples and the reference sample. This sampling and recurrent ranking method modifies the statistics of the considered tests, so one needs to evaluate new statistics distributions. It is important to note, that this sampling method can be applied to other rank tests.

A coarse approximation with normal distribution can be used for the joint test with independent statistics distributions. Results that are more interesting can be obtained using logical conjunction of decisions, because this operation allows to independently set the detection threshold for a given false alarm probability criteria.

Keywords: nonparametric test, sampling method, digital signal processing.

Acknowledgements. Research has been carried out with financial support from Ministry of Education and Science of Russian Federation, state assignment № 2014/138, project № 1176

Received 5 November 2015