

ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

УДК 517.518.8

Ю. Е. ВОСКОБОЙНИКОВ

(Новосибирск)

ПОСТРОЕНИЕ ДЕСКРИПТИВНЫХ ПРИБЛИЖЕНИЙ ДЛЯ СГЛАЖИВАНИЯ И ДИФФЕРЕНЦИРОВАНИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Постановка задачи. Часто результаты экспериментального исследования функциональной зависимости $y(x)$ представлены таблицей

$$x_i, \tilde{y}_i, \quad i = 1, \dots, n, \quad (1)$$

где $\tilde{y}_i = y(x_i) + \xi_i$ — измеренные значения функции, ξ_i — шум i -го измерения. Типовыми задачами обработки таких данных являются сглаживание (уменьшение уровня шума) и вычисление производных требуемого порядка. Последняя задача, как известно, относится к некорректно поставленным [1]. Большинство методов устойчивого решения этих задач основано на приближении (аппроксимации) функции $y(x)$ полиномами заданной степени (например, сглаживающими кубическими сплайнами [2, 3]).

В данной работе приближение таблично заданной функции строится в базисе B -сплайнов. При вычислении коэффициентов приближения в этом базисе может существенно использоваться априорная информация о функции $y(x)$, вводимая системой неравенств

$$y^{(l_i)}(x_i^*) \begin{cases} \leq \\ = \\ \geq \end{cases} d_i, \quad i = 1, \dots, N_r, \quad (2)$$

где $y^{(l_i)}(x_i^*)$ — производная l_i -го порядка функции $y(x)$ в точке x_i^* ,

а обозначение $\begin{cases} \leq \\ = \\ \geq \end{cases}$ указывает, что в качестве отношения может выступать любое из трех отношений, указанных в скобках. Ограничения (2) часто носят качественный, описательный характер (т. е. больше или меньше заданной величины). Поэтому приближение функции $y(x)$, удовлетворяющее ограничениям (2), будем называть дескриптивным. В отличие от работы [4], в которой строились только выпуклые среднеквадратичные приближения, учет ограничений дает возможность более полно использовать априорную информацию, т. е. строить приближения с различными заданными свойствами (монотонность, выпуклость, вогнутость, требуемые краевые условия, наличие экстремумов в заданных точках и т. д.).

Определение пространства приближающих функций. Рассмотрим вопрос о выборе пространства, элементами которого будет осуществляться аппроксимация функции $y(x)$, заданной таблицей (1). Представляется удобным в качестве такого пространства взять множество $P_{m, z, \theta}$ кусочных полиномов m -й степени [5]. Для определения пространства $P_{m, z, \theta}$ введем две последовательности: строго возрастающую последовательность

действительных чисел $z = \{z_1, z_2, \dots, z_{N_z}\}$, при этом $z_1 \leq \min(x_1, x_2, \dots, x_n)$, $z_{N_z} \geq \max(x_1, x_2, \dots, x_n)$, и последовательность положительных целых чисел $\theta = \{\theta_1, \theta_2, \dots, \theta_{N_z}\}$. Тогда, если функция $\varphi(x) \in P_{m, z, \theta}$, то

$$\varphi(x) = \begin{cases} P_1(x), & x < z_1; \\ P_i(x), & z_i \leq x \leq z_{i+1}; \\ P_{N_z}(x), & x > z_{N_z} \end{cases}$$

$$\text{и } \text{jumpr}_{z_i} \left(\frac{d^k \varphi(x)}{dx^k} \right) = 0, \quad k = 0, \dots, \theta_i - 1, \quad i = 2, \dots, N_z - 1,$$

где $P_i(x)$ — полиномы степени m с действительными коэффициентами; $\text{jumpr}_{z_i} \varphi(x) = \varphi(z_i^+) - \varphi(z_i^-)$ — скачок функции $\varphi(x)$ в точке $x = z_i$. Аппроксимация таблично заданной функции элементами пространства $P_{m, z, \theta}$ позволяет получать хорошее приближение функции $y(x)$ (соответствующим выбором степени m и последовательности z) с требуемыми свойствами непрерывности. Так, приняв $\theta_i = m$, $i = 2, \dots, N_z - 1$, получаем приближение с непрерывными на интервале (z_1, z_{N_z}) производными до $(m-1)$ -й включительно, являющееся сплайном степени m дефекта 1 [6]. При необходимости, положив $\theta_i = 0$, можно найти приближение с разрывом 1-го рода в точке $x = z_i$.

Удобным базисом пространства $P_{m, z, \theta}$ может служить последовательность нормализованных B -сплайнов (базисных сплайнов) степени m [5, 6]. Если узлы $\{\tilde{x}_i\}_{i=1}^{N+m+1}$ B -сплайнов определены через элементы z_i соотношениями

$$\begin{aligned} \tilde{x}_1 &\leq \tilde{x}_2 \leq \dots \leq \tilde{x}_{m+1} = z_1; \\ \tilde{x}_{l(j)} &= \tilde{x}_{l(j)+1} = \dots = \tilde{x}_{l(j+1)-1} = z_j, \quad j = 2, \dots, N_z - 1; \\ z_{N_z} &= \tilde{x}_{N+1} \leq \tilde{x}_{N+2} \leq \dots \leq \tilde{x}_{N+m+1}, \end{aligned} \quad (3)$$

$$\begin{aligned} \text{где } l(j) &= m + 1 + \sum_{i=2}^j (m + 1 - \theta_i), \\ N &= m + 1 + \sum_{i=2}^{N_z-1} (m + 1 - \theta_i), \end{aligned} \quad (4)$$

то справедливо следующее утверждение [5]:

Последовательность нормализованных B -сплайнов $B_{m, 1}, B_{m, 2}, \dots, B_{m, N}$ m -й степени с узлами (3) есть базис пространства $P_{m, z, \theta}$ на интервале (z_1, z_{N_z}) .

Следовательно, искомое приближение функции $y(x)$ элементом пространства $P_{m, z, \theta}$ может быть представлено комбинацией

$$S_N(x, \mathbf{a}) = \sum_{j=1}^N a_j B_{m, j}(x). \quad (5)$$

Такое приближение полностью определяется коэффициентами a_j , $j = 1, \dots, N$, представляющими собой координаты элемента пространства $P_{m, z, \theta}$ в базисе B -сплайнов. По найденным коэффициентам a_j значения приближения и его производных на сетке с узлами $x_i \in (z_1, z_{N_z})$ определяются выражением

$$S_N^{(k)}(x_i, \mathbf{a}) = \sum_{j=1}^N a_j B_{m, j}^{(k)}(x_i). \quad (6)$$

Алгоритм вычисления коэффициентов дескриптивного приближения. При заданных параметрах m, z, θ пространства $P_{m, z, \theta}$ вычисление вектора коэффициентов $\mathbf{a} = \{a_1, a_2, \dots, a_N\}$ целесообразно осуществлять из

условия минимума функционала

$$\Phi(\mathbf{a}) = \sum_{i=1}^n p_i (\tilde{y}_i - S_N(x_i, \mathbf{a}))^2$$

при ограничениях (2). Весовые множители p_i строго положительны и характеризуют значимость i -го измерения. Введя в рассмотрение симметричную матрицу \mathbf{Q} с элементами

$$Q_{kj} = \sum_{i=1}^n p_i B_{m,k}(x_i) B_{m,j}(x_i)$$

и вектор \mathbf{g} с проекциями

$$q_j = \sum_{i=1}^n p_i \tilde{y}_i B_{m,j}(x_i),$$

функционал $\Phi(\mathbf{a})$ удастся представить в виде квадратичного функционала

$$\Phi(\mathbf{a}) = \mathbf{a}^T \mathbf{Q} \mathbf{a} - 2 \mathbf{a}^T \mathbf{q} + \sum_{i=1}^n p_i \tilde{y}_i^2. \quad (7)$$

Обозначив $c_{ij} = B_{m,j}^{(l_i)}(x_i^*)$, ограничения (2) можно записать в виде системы неравенств

$$\begin{aligned} g_i(\mathbf{a}) &= \sum_{j=1}^N c_{ij} a_j - d_i = 0, \quad i = 1, \dots, n_r; \\ g_i(\mathbf{a}) &= \delta_i \left(\sum_{j=1}^N c_{ij} a_j - d_i \right) \leq 0, \quad i = n_r + 1, \dots, N_r, \end{aligned} \quad (8)$$

где n_r — число ограничений, имеющих вид равенств, множитель δ_i принимает значение 1, если i -е ограничение имеет вид \leq , и $\delta_i = -1$, если \geq . Эта система определяет в N -мерном пространстве коэффициентов a_j некоторую допустимую область Ω_d , точки которой удовлетворяют (8). Таким образом, задача определения коэффициентов дескриптивного приближения свелась к задаче квадратичного программирования: найти вектор коэффициентов $\mathbf{a}^{**} \in \Omega_d$, доставляющий минимум (7), т. е.

$$\Phi(\mathbf{a}^{**}) = \inf_{\mathbf{a} \in \Omega_d} \Phi(\mathbf{a}). \quad (9)$$

Естественным предположением при этом является допущение о совместности ограничений (2), т. е. область Ω_d должна содержать хотя бы одну точку. В противном случае построение дескриптивного приближения теряет смысл.

Для решения задачи квадратичного программирования использовался метод штрафных функций. Не излагая самого метода [7], заметим только, что основная его идея заключена в приближенном сведении задачи (9) к последовательности задач минимизации без ограничений некоторого вспомогательного функционала. В качестве последнего принят

$$F(\mathbf{a}, \gamma) = \Phi(\mathbf{a}) + \gamma G(\mathbf{a}), \quad (10)$$

где γ — штрафной множитель, $G(\mathbf{a})$ — штрафной функционал вида

$$G(\mathbf{a}) = \sum_{i=1}^{n_r} \omega_i |g_i(\mathbf{a})|^2 + \sum_{i=n_r+1}^{N_r} \omega_i [\max(0, g_i(\mathbf{a}))]^2,$$

который строго больше нуля, если \mathbf{a} не принадлежит допустимой области Ω_d , и равен нулю при $\mathbf{a} \in \Omega_d$. Для функционала (10) справедливы два утверждения [7].

Минимум функционала $F(\mathbf{a}, \gamma)$ при любой величине $\gamma \in (0, \infty)$ достигается в единственной точке \mathbf{a}_γ^* .

Если последовательность \mathbf{a}_γ^* определяется из условия минимума (10) при $\gamma \rightarrow \infty$, то предельная точка этой последовательности \mathbf{a}_∞^* есть решение задачи (9), а значения $\Phi(\mathbf{a}_\gamma^*)$ монотонно возрастают и $\lim_{\gamma \rightarrow \infty} \Phi(\mathbf{a}_\gamma^*) = \Phi(\mathbf{a}^{**})$.

В качестве приближенного решения задачи (9) принимается элемент \mathbf{a}_γ^{**} последовательности \mathbf{a}_γ^* , удовлетворяющий условиям

$$\begin{aligned} \omega_i |g_i(\mathbf{a}_\gamma^{**})| &\leq \varepsilon_1, & i = 1, \dots, n_r; \\ \omega_i (g_i(\mathbf{a}_\gamma^{**})) &\leq \varepsilon_2, & i = n_r + 1, \dots, N_r, \end{aligned}$$

где $\varepsilon_1, \varepsilon_2$ — положительные константы, определяющие (с учетом весовых множителей ω_i) точность выполнения ограничений (8). Вышеизложенное обобщает следующее утверждение:

При любых значениях (1), фиксированных параметрах m, z, θ и совместных ограничениях (2) существует единственное дескриптивное приближение таблично заданной функции элементом (5) пространства $P_{m, z, \theta}$.

О выборе размерности пространства $P_{m, z, \theta}$. В предыдущем разделе при определении коэффициентов дескриптивного приближения $S_N(x, \mathbf{a})$ предполагалось, что размерность N пространства $P_{m, z, \theta}$ задана и фиксирована. Заметим, что размерность N является своеобразным параметром сглаживания, который присутствует в сглаживающих кубических сплайнах [2, 3]. Действительно, при увеличении N введением новых значений z_i (и соответствующих θ_i) появляются новые базисные функции, что дает возможность повысить точность приближения функции $y(x)$ комбинацией (5). Однако размерность пространства $P_{m, z, \theta}$ должна согласовываться с уровнем шума измерения. При отсутствии такого согласования можно получить либо «переглаженное» приближение (заниженное значение N), либо приближение, в первой производной которого присутствуют шумовые осцилляции (завышенная размерность пространства). Следовательно, необходимо определить оптимальную размерность N_{opt} , которая минимизирует среднеквадратическую ошибку приближения функции $y(x)$ комбинацией (5) среди всех других значений N . Ниже приводятся два метода оценивания величины N_{opt} при различной априорной информации о шуме измерения.

Метод 1. Вектор шума измерения $\xi = \{\xi_1, \xi_2, \dots, \xi_n\}$ имеет нулевое среднее и известную диагональную корреляционную матрицу $V_\xi = \text{diag} \{\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2\}$. Введем в рассмотрение билинейную форму

$$R(N) = \sum_{i=1}^n \tilde{y}_i e_i(N) / \sigma_i^2.$$

Величина $e_i(N) = \tilde{y}_i - S_N(x_i, \mathbf{a}^{**})$ — невязка i -го измерения. В качестве оценки N_{opt} принимается величина N_R , при которой значение $R(N_R)$ попадает в интервал $[\Theta_M(\beta/2), \Theta_M(1 - \beta/2)]$, где $\Theta_M(\beta/2)$ — квантиль порядка $\beta/2$ χ_M^2 -распределения. Число степеней свободы M определяется выражением

$$M = n - N + n_a, \quad (11)$$

в котором n_a — количество ограничений системы (8), обращающихся при вычисленных значениях \mathbf{a}^{**} в тождества с заданной точностью. Оценка N_R оптимальной размерности пространства $P_{m, z, \theta}$ следует из критерия оптимальности приближения экспериментальной информации [8].

Метод 2. Вектор ξ имеет нулевое среднее и неизвестную корреляционную матрицу. В этом случае оценкой N_{opt} является значение N_V , доставляющее минимум функционалу

$$V(N) = \frac{1}{n} \sum e_i^2(N) \left/ \left[1 - \frac{N - n_a}{n} \right] \right.,$$

т. е.

$$V(N_V) = \inf_{N > 0} V(N). \quad (12)$$

Этот метод — обобщение метода «cross — validation» [9] оценивания параметра сглаживания на случай, когда искомое приближение имеет вид (5).

Программная реализация и результаты вычислительного эксперимента. Изложенный алгоритм построения дескриптивного приближения реализован в виде комплекса подпрограмм, написанных на языке ФОРТРАН-IV [10]. В комплексе

предусмотрено обнаружение и диагностика ошибок в исходных данных, а также ошибок, возникающих в ходе вычислительного процесса построения дескриптивного приближения. По функциональному назначению модули комплекса можно разделить на следующие группы:

модули, обнаруживающие и диагностирующие ошибки;

модули решения задачи квадратичного программирования;

модули вычисления B -сплайнов;

модули определения оптимальной размерности пространства $P_{m, z, o}$.

Остановимся на некоторых результатах вычислительного эксперимента с описанным комплексом подпрограмм.

На интервале $[a_x, b_x]$ задавалась функция $y(x) = \exp(-(x-2)^2/8)$, значения которой вычислялись в узлах $x_i = a_x + (i-1)\Delta_x$, $i = 1, \dots, n$, где $\Delta_x = (b_x - a_x)/(n-1)$, $a_x = 1$, $b_x = 8$, $n = 40$. Вычисленные значения $y(x_i)$ искажались нормально распределенными случайными величинами ξ_i с нулевым средним и дисперсией $\sigma_i^2 = 0,01$, что примерно соответствует относительному уровню шума измерений в 25%. По исходным данным $\{x_i, \tilde{y}_i, p_i = 1/\sigma_i^2\}$ строились два приближения:

среднеквадратическое приближение $S_N(x, a^*) = \sum_{j=1}^N a_j^* B_{m,j}$, коэффициенты которого определялись из условия минимума функционала (7) без учета ограничений (2);

дескриптивное приближение $S_N(x, a^{**})$, коэффициенты которого определялись из решения задачи квадратичного программирования (9) при ограничениях, приведенных в табл. 1.

Точность приближения функции $y(x)$ и ее производной $y'(x)$ эле-

Таблица 1

Номер ограничения	x_i^*	l_i	d_i	Знак отношения
1	1,000	1	0,1174	=
2	1,359	1	0,0000	>
3	1,897	1	0,0000	>
4	2,436	1	-0,0500	>
5	4,590	1	-0,1500	>
6	7,462	1	-0,0500	>
7	8,000	1	-0,0790	=

Таблица 2

Размерность N	Элементы z_i	Приближение $S_N(x, a^*)$		Приближение $S_N(x, a^{**})$	
		$\varepsilon_y(a^*)$	$\varepsilon_{y'}(a^*)$	$\varepsilon_y(a^{**})$	$\varepsilon_{y'}(a^{**})$
4	1; 9	0,0983	0,0934	0,1015	0,0941
5	1; 3; 9	0,2144	0,5558	0,1101	0,1003
6	1; 2; 4; 9	0,2537	1,2675	0,1061	0,0858
7	1; 2; 3; 5; 9	0,2699	1,4587	0,0877	0,0845
8	1; 2; 3; 5; 7; 9	0,3916	1,9575	0,0850	0,0855
9	1; 1,5; 2,5; 3; 5; 7; 9	0,3928	1,9581	0,0921	0,1121
10	1; 1,5; 2; 2,5; 3; 5; 7; 9	0,3917	1,9599	0,1129	0,1341

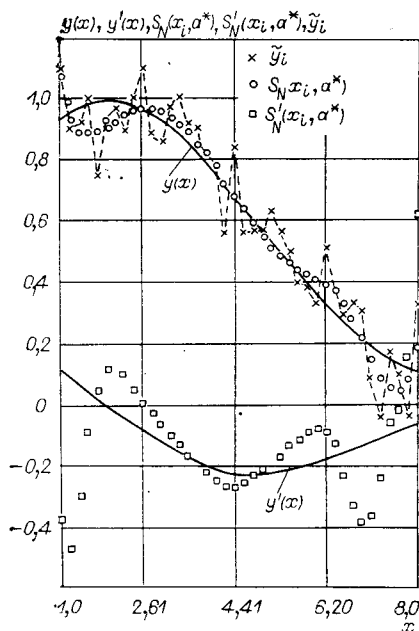


Рис. 1.

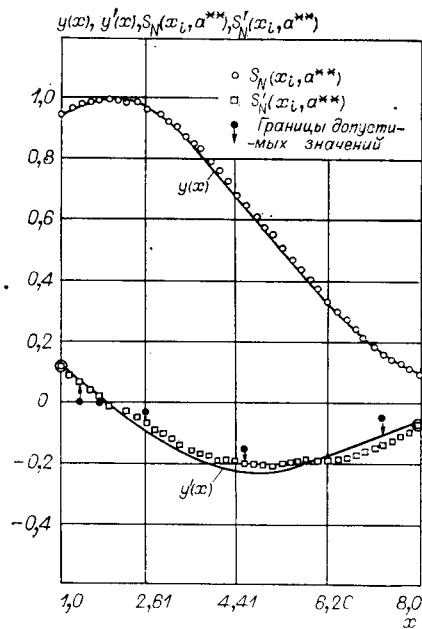


Рис. 2.

ментами пространства $P_{m, z, \theta}$ различной размерности N характеризовались среднеквадратическими ошибками

$$\varepsilon_y(\mathbf{a}) = \left[\sum_{i=1}^n (y(x_i) - S_N(x_i, \mathbf{a}))^2 \right]^{1/2},$$

$$\varepsilon_{y'}(\mathbf{a}) = \left[\sum_{i=1}^n (y'(x_i) - S'_N(x_i, \mathbf{a}))^2 \right]^{1/2}.$$

Результаты приведены в табл. 2. На рис. 1, 2 представлены функции $y(x)$, $y'(x)$ и соответствующие значения приближений $S_N(x, \mathbf{a}^*)$, $S'_N(x, \mathbf{a}^{**})$. На рис. 2 также показаны точки, определяющие допустимые значения первой производной в узлах $[x_i^*]$ (стрелки направлены внутрь допустимой области). Анализ этого вычислительного эксперимента, а также обработка реальных экспериментальных данных показали, что построение дескриптивных приближений существенно (на порядок и более) повышает точность дифференцирования таблично заданной функции.

Важным преимуществом дескриптивного приближения по сравнению со среднеквадратическим является слабая зависимость ошибок сглаживания и дифференцирования от размерности пространства $P_{m, z, \theta}$ (см. табл. 2). Это объясняется тем, что учет априорной информации оказывает более сильное регуляризующее воздействие на построенное приближение, чем изменение размерности пространства.

В заключение заметим, что изложенный подход может быть использован для построения дескриптивных решений интегральных уравнений 1-го рода соответствующим изменением матрицы Q и вектора q функционала (7).

ЛИТЕРАТУРА

1. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач.— М.: Наука, 1979.
2. Стечкин С. В., Субботин Ю. И. Сплайны в вычислительной математике.— М.: Наука, 1976.
3. Воскобойников Ю. Е. Построение сглаживающих сплайнов при машинной обработке результатов эксперимента.— Автометрия, 1979, № 4.
4. Dierckx P. Algorithm 42. An algorithm for cubic spline fitting with convexity constraints.— Computing, 1980, vol. 34, N 4, p. 449—374.
5. Boor C. A practical guide to spline functions.— N. Y.: Springer Verlag, 1978.

6. Завьялов Ю. С., Квасов Б. И., Мирошниченко В. Л. Методы сплайн-функций.— М.: Наука, 1980.
7. Васильев Ф. П. Численные методы решения экстремальных задач.— М.: Наука, 1980.
8. Воскобойников Ю. Е. Критерий и алгоритмы выбора параметра при сглаживании сплайн-функциями.— В кн.: Алгоритмы обработки и средства автоматизации теплофизического эксперимента. Новосибирск: ИТФ СО АН СССР, 1978.
9. Golub H. G., Heath M. Generalized cross-validation as a method for choosing a good ridge parameter.— *Technometrics*, 1979, vol. 21, N 2, p. 215—223.
10. Воскобойников Ю. Е. Комплекс программ для сглаживания и дифференцирования экспериментальных данных при помощи В-сплайнов.— В кн.: Алгоритмические и аппаратные средства переработки информации. Новосибирск: ИТФ СО АН СССР, 1981.

Поступила в редакцию 6 января 1982 г.

УДК 621.391.1 : 519.246

Н. Г. ЧЕРНОГУЗ

(Баку)

ОТБРАКОВКА АНОМАЛЬНЫХ ОШИБОК С ПОМОЩЬЮ СТОХАСТИЧЕСКИХ МОДЕЛЕЙ

Существующие подходы к анализу применяемых в информационно-измерительных системах разностных алгоритмов отбраковки аномальных ошибок [1, 2] допускают определенное упрощение задачи и поэтому не дают достаточно полного представления об их эффективности. Целью настоящей работы является анализ алгоритма отбраковки аномальных ошибок при измерении центрированного нормального марковского случайного процесса $\{y_n\}$, $n = 0, 1, 2, \dots$, описываемого стохастическим разностным уравнением авторегрессии 1-го порядка [3]:

$$y_n = \rho y_{n-1} + v_n, \quad (1)$$

где $\rho = E[y_n y_{n-1}] / \sigma_y^2$, $\sigma_y^2 = D[y_n]$ ($E[\]$ — знак математического ожидания, $D[\]$ — дисперсии), $\{v_n\}$ — «белый» гауссов шум.

Рассмотрим аномальные ошибки как стационарный пуассоновский поток импульсов, имеющих величину λ и среднюю частоту ν , причем при появлении аномальной ошибки результат измерения x_n принимает значение $y'_n = y_n + \lambda$, а при ее отсутствии $x_n = y_n$.

Алгоритм отбраковки аномальных ошибок заключается в том, что для проверяемого в текущий момент наблюдения x_n вычисляется прогноз $\hat{y}_n = \rho z_{n-1}$, где z_{n-1} — результат проверки предыдущего наблюдения (причем в общем случае $\rho \neq 1$); и если абсолютная величина ошибки прогноза $u_n = x_n - \hat{y}_n$ превышает некоторый допустимый порог, то проверяемое наблюдение x_n заменяется своим прогнозом, т. е. $z_n = \rho z_{n-1}$, в противном случае $z_n = x_n$.

Пользуясь установившейся для задачи отбраковки терминологией [2], рассмотрим процесс принятия решения при проверке наблюдения x_n как проверку гипотезы $H_0: x_n = y'_n$ против альтернативы $H_1: x_n = y_n$. Ошибка в принятии решения при этом характеризуется двумя условными вероятностями: вероятностью пропуска аномальной ошибки $P(H_1|H_0)$ и вероятностью ложной отбраковки $P(H_0|H_1)$. Качество алгоритма отбраковки можно характеризовать, например, полной вероятностью ошибочных решений [4]

$$L = P(H_1|H_0)P_{H_0} + P(H_0|H_1)P_{H_1}, \quad (2)$$