2.9. Знаковый обнаружитель изменения медианы распределения

Рассмотрим задачу обнаружения изменения распределения независимой однородной выборки $y = (y_1, y_2, \ldots, y_N)$, что соответствует появлению сигнала на фоне стационарного шума.

Предположим, что шум имеет плотность вероятности $f_0(y)$ с известной медианой $y_{\text{med}0}$. В остальном форма плотности вероятности шума может быть произвольной. Это соответствует сложной гипотезе и *непараметрической модели* для плотности шума, поскольку класс плотностей с заданной медианой не может быть аналитически описан в параметрической форме.

Задача состоит в проверке гипотезы $H_0 : y_{\text{med}} = y_{\text{med}0}$ о медиане y_{med} плотности вероятности $f(y)$ наблюдений против сложной альтернативы $H_1 : y_{\text{med}} \neq y_{\text{med}0}$, т. е. изменение медианы связано с появлением полезного сигнала. Обозначим $p = P(y_k > y_{\text{med}0})$ – вероятность превышения выборкой медианы. Тогда, если сигнала нет, $p_0 = 1/2$, а при появлении сигнала $p_1 \neq 1/2$. Фактически задача формулируется как обнаружение изменения медианы распределения.

Решение задачи возможно с привлечением знаковых и ранговых статистик, которые обладают *непараметрическим свойством*, т. е. обеспечивают стабилизацию вероятности ложной тревоги при изменениях вида и параметров плотности вероятности шума.

Знаковая статистика $u = \sum_{k=1}^{N} \operatorname{sgn}(y_k - y_{\text{med}0})$ принимает целочисленные значения от $-N$ до N и имеет при отсутствии сигнала симметричное распределение относительно нуля. Можно использовать *односторонний знаковый* алгоритм $u \geq u_T$ для обнаружения положительных изменений медианы или *двусторонний знаковый* алгоритм $|u| \geq u_{T2}$ для обнаружения изменений разных знаков.

Эквивалентная решающая статистика $n = \sum_{k=1}^{N} \operatorname{step}(y_k - y_{\text{med}0})$ принимает целочисленные значения от нуля до N. Обе статистики используют *бинарное квантование* с порогом квантования, равным медиане.

Функция единичного скачка $\operatorname{step}(\cdot)$ фактически осуществляет бинарное квантование исходной выборки на уровни нуля и единицы. Решающая статистика n формируется на выходе счетчика полученных единиц, реализуя *метод бинарного интегрирования* [14].

Структура знакового обнаружителя положительных изменений медианы представлена на рис. 2.49.

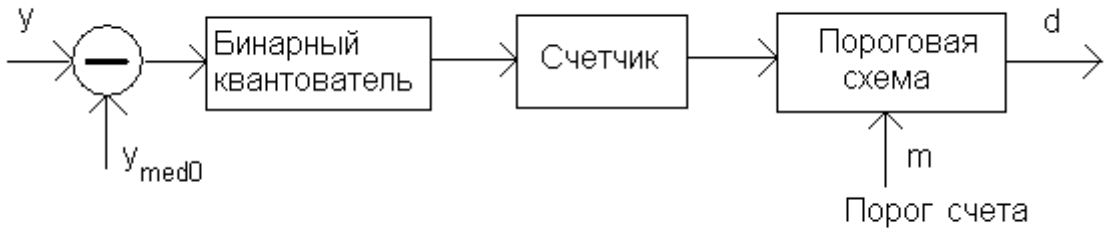

Рис. 2.49. Структура знакового обнаружителя положительного изменения медианы

Знаковый алгоритм обнаружения выбирает решение d_1, если $n \geq m$, где m есть односторонний целочисленный порог счета.

В общем виде алгоритм обнаружения известен как тест «m из N» [14] и может использовать вместо медианы произвольный порог y_0 бинарного квантования: $n = \sum_{k=1}^{N} \mathrm{step}(y_k - y_0) \geq m$.

Статистика n при обеих гипотезах имеет биномиальное распределение с параметрами N и p. Вероятность достижения или превышения порога в общем случае равна $P(n \geq m) = \sum_{k=m}^{N} C_N^k p^k (1-p)^{N-k} = \mathrm{I}_p(m, N-m+1)$. Здесь $\mathrm{I}_x(a,b)$ представляет собой неполную бета-функцию [1], а C_N^k – биномиальные коэффициенты.

При гипотезе H_0: $p_0 = 1/2$ вероятность ложной тревоги $F = (1/2^N) \sum_{k=m}^{N} C_N^k$ не зависит от распределения шума, что отражает *непараметрическое свойство* обнаружителя стабилизировать вероятность ложной тревоги [1].

При наличии полезного сигнала H_1: $p_1 > 1/2$ и качество обнаружения зависит от относительного изменения $\delta p = (p_1 - p_0)/p_0$, при этом δp меняется от нуля до единицы. Другим параметром обнаружения может служить введенное в [14] *выходное отношение сигнал/шум* $\rho_N = \sqrt{N}(p_1 - p_0)/\sqrt{p_0(1-p_0)}$, которое связано с δp

следующим выражением $\rho_N = \sqrt{Np_0/(1-p_0)}\delta p$, и представляет собой дефлекцию решающей статистики n.

При больших размерах выборки можно использовать гауссовскую аппроксимацию $N(\mu_a, \sigma_a^2)$ биномиального распределения, которое имеет математическое ожидание $\mu_a = Np$ и дисперсию $\sigma_a^2 = Np(1-p)$. При гипотезе H_0 имеем $\mu_{a0} = N/2$, $\sigma_{a0}^2 = N/4$, а при гипотезе H_1 увеличивается как математическое ожидание, так и дисперсия распределения $\mu_{a1} = Np_1$, $\sigma_{a1}^2 = Np_1(1-p_1)$.

Структура знакового двухальтернативного обнаружителя (различителя) изменений обоих знаков представлена на рис. 2.50 и использует два порога счета. Положительное изменение регистрируется при $n \geq m^+$, а отрицательное – при $n \leq m^-$. В остальных случаях принимается решение о совпадении медианы с заданным значением y_{med0}.

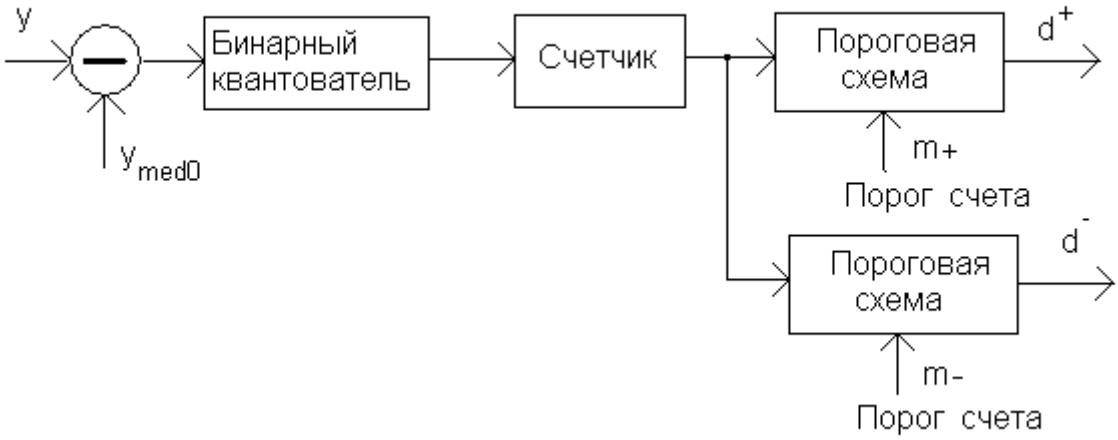

Рис. 2.50. Структура знакового различителя изменения медианы распределения

Пример. Рассмотрим случай $N = 81$, для него из таблиц биномиального распределения получаем $P_0(n \geq 51) = 0{,}0128$, $P_0(n \geq 52) = 0{,}007$. Для критерия Неймана–Пирсона выполнение условия на вероятность ложной тревоги $F \leq 0{,}01$ дает пороговое значение $m_{NP} = 52$. Аналогично для $F \leq 0{,}001$ имеем $P_0(n \geq 54) = 0{,}0018$ и $P_0(n \geq 55) = 0{,}00084$. Зависимость F от дискретного порога m представлена на рис. 2.51. Гауссовская аппроксимация (штриховая линия) дает слегка заниженные результат $P_{0A}(n \geq 51) = 0{,}0098$, $P_{0A}(n \geq 52) = 0{,}0053$ и $P_{0A}(n \geq 54) = 0{,}00135$, $P_{0A}(n \geq 55) = 0{,}00064$.

Math_doc\Beta.mcd

× ×

Рис. 2.51. Зависимость вероятности ложной тревоги от дискретного порога для $N = 81$

Характеристики обнаружения в зависимости от p_1 приведены на рис. 2.52 (соответствующие кривые помечены крестиками). Зависимость 1 соответствует порогу $m = 47$ и вероятности ложной тревоги $P_0(n \geq 47) = 0{,}09$, зависимость 2 построена для порога $m = 52$, $P_0(n \geq 52) = 0{,}007$, а зависимость 3 – для $m = 55$, $P_0(n \geq 55) = 0{,}00084$. Соответствующие гауссовские аппроксимации дают несколько худшие характеристики и представлены штриховыми линиями.

Можно определить пороговые значения для p_1 и δp, если задать уровни вероятности правильного обнаружения D. В частности, для порога $m = 52$ пороговые значения $p_1^{(0,5)} = 0{,}635$ и $p_1^{(0,9)} = 0{,}7$ обеспечивают $D = 0{,}5$ и $D = 0{,}9$ соответственно.

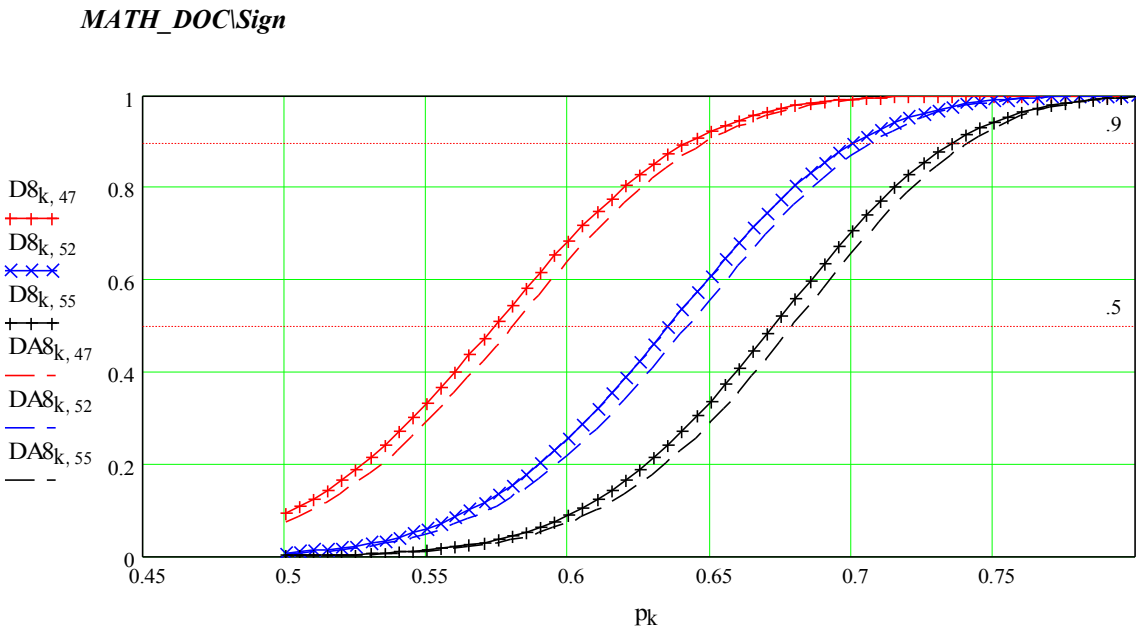

Рис. 2.52. Характеристики обнаружения знакового обнаружителя с одним порогом для $N = 81$

Для байесовского критерия можно указать оптимальное значение порога m_{opt}, которое обеспечивает минимум суммарной вероятности ошибки $p_\Sigma = (F + 1 - D)/2$ для $p_0 = 0{,}5$. На рис. 2.53 представлены вероятности p_Σ в зависимости от порога m (характеристики существуют только в дискретных точках). Линии 1–4 соответствуют значениям $p_1 = 0{,}635;\ 0{,}7;\ 0{,}8;\ 0{,}9$. Оптимальные значения m_{opt} меняются от 47 до 60.

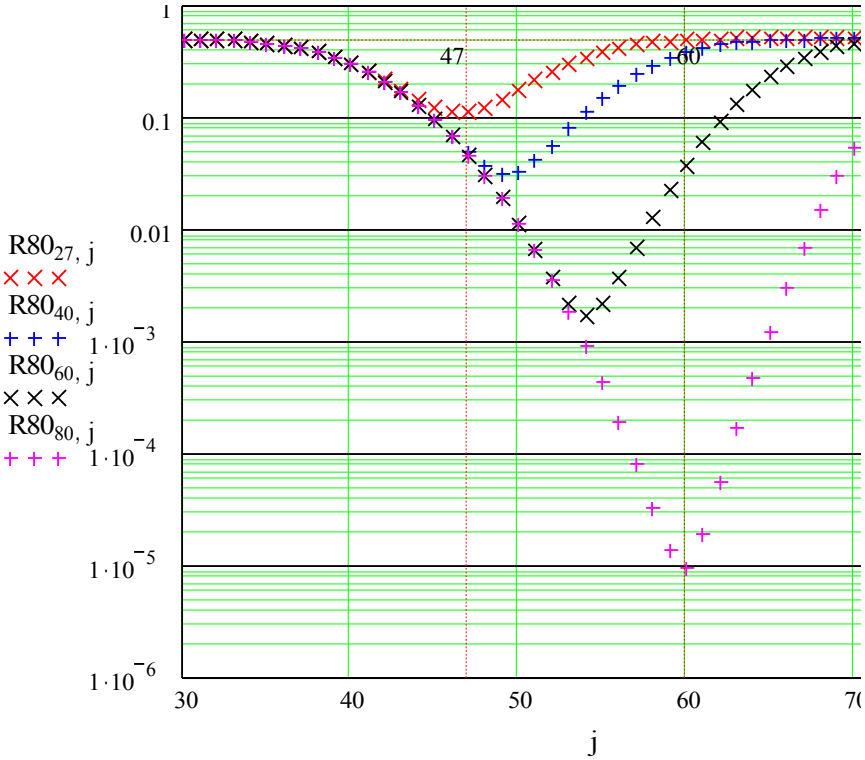
1
0.1
0.01
1·10⁻³
1·10⁻⁴
1·10⁻⁵
1·10⁻⁶
47
60
R80₂₇, j
R80₄₀, j
R80₆₀, j
R80₈₀, j
30
40
50
60
70
j

Рис. 2.53. Зависимости суммарной вероятности ошибки от значений порога для $p_0 = 0{,}5$

Характеристики обнаружения
для гауссовских наблюдений $y_1, y_2, \dots, y_N$.

Пусть $f_0(y_k) \sim N(m_0, \sigma_0^2)$ и $f_1(y_k) \sim N(m_1, \sigma_0^2)$. Медианы y_{med0} и y_{med1} совпадают с математическими ожиданиями m_0 и m_1. В качестве параметра обнаружения можно использовать входное отношение сигнал/шум $d = (m_1 - m_0)/\sigma_0$. Фактически задача сводится к обнаружению сдвига гауссовского распределения. Вероятность превышения порога при появлении сигнала равна $p_1 = \Phi(d)$, где $\Phi(z) = (1/\sqrt{2\pi})\int_{-\infty}^{z} \exp(-t^2/2)\mathrm{d}t$ – интеграл вероятности Лапласа.

Если присутствует полезный сигнал с отношением сигнал/шум $d = 0{,}258$, то $p_1 = 0{,}602$, в этом случае $\delta p = 0{,}2$. Бинарное накопление при $N = 81$ и сравнение с порогом $m = 52$ дает $P_1(n \geq 52) = 0{,}269$. Рабочие характеристики приемника существуют только в отдельных точках на плоскости (D, F) [15, 16]. Гауссовская аппроксимация здесь также дает заниженные значения $P_{1A}(n \geq 52) = 0{,}231$.

Пороги для различения сдвигов, имеющих разные знаки, равны $n^{+}_{NP} = 52$ и $n^{-}_{NP} = 28$, они обеспечивают $F \approx 0{,}01$. Вероятность правильного обнаружения $D = 0{,}269$, при этом гауссовская аппроксимация дает значение $D = 0{,}231$.

Заметим, что этот анализ справедлив для любой плотности шума $f_0(y)$ с известной медианой, что является следствием непараметрического свойства.

В случае неизвестной медианы распределения шума используются *двухвыборочные алгоритмы* с опорной выборкой шума $x = (x_1, x_2, ..., x_N)$. Проверяется гипотеза о равенстве медиан анализируемой и опорной выборок.

Например, алгоритм $n = \sum_{k=1}^{N} \mathrm{step}(y_k - \hat{x}_{\mathrm{med}}) \geq m$ использует оценку медианы в выборке x вместо неизвестного значения медианы y_{med}. Если известно, что распределение шума симметрично, то вместо оценки медианы можно взять выборочное среднее $\bar{x}$, которое имеет меньшую дисперсию. Эти алгоритмы должны обладать характеристиками обнаружения, близкими к одновыборочному.

Двухвыборочный знаковый алгоритм $n = \sum_{k=1}^{N} \mathrm{step}(y_k - x_k) \geq m$ является *тестом согласия* и проверяет более общую гипотезу об одинаковости распределений двух выборок. Вследствие этого потери возрастают. Вероятность правильного обнаружения зависит теперь от величины $p_1 = P_1(y > x) = \Phi(d/2)$. При $d = 0{,}258$ имеем $p_1 = 0{,}551$ и бинарное накопление при $N = 81$ с порогом $m = 52$ дает только $P_1(n \geq 52) = 0{,}062$.

Характеристики обнаружения знаковых обнаружителей показаны на рис. 2.54 (линии отмечены крестиками). Пороговые дефлекции двухвыборочного алгоритма (правая кривая) увеличиваются вдвое по сравнению с одновыборочным (средняя кривая). И тот и другой

проигрывают линейному накопителю, характеристика обнаружения которого показана штриховой линией (левая кривая).

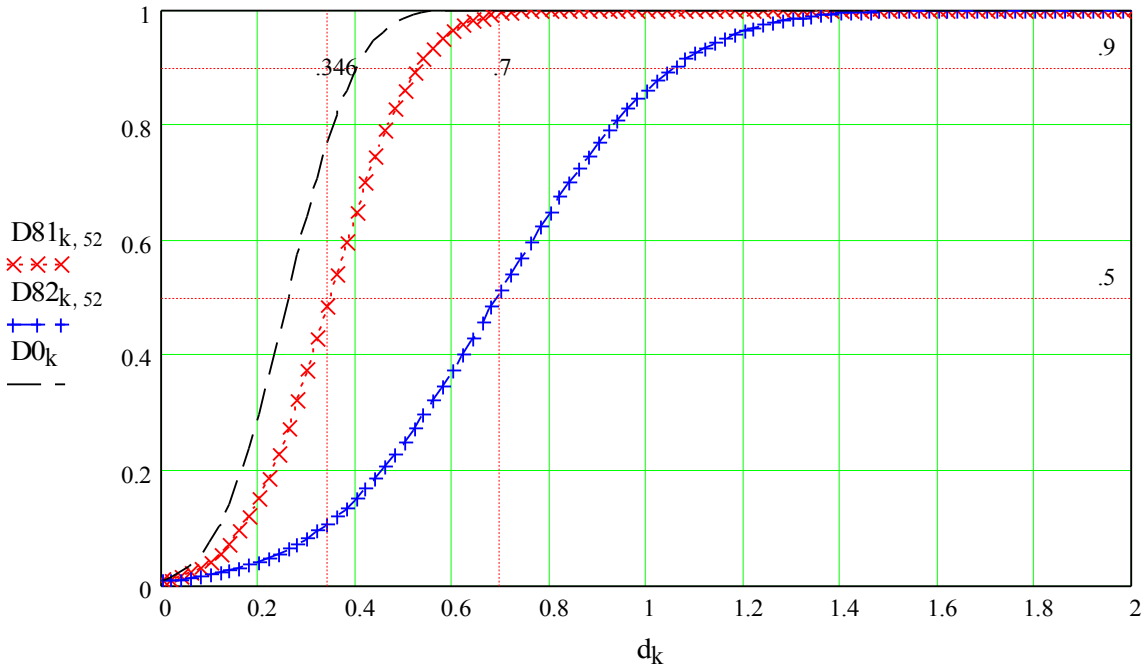

Рис. 2.54. Характеристики обнаружения знаковых обнаружителей сдвига гауссовского распределения **MATH_DOC\Sign**

Если проигрыш одновыборочного знакового алгоритма не так значителен, то двухвыборочный знаковый тест согласия уже значительно уступает линейному накопителю. Эти потери представляют собой плату за инвариантность знаковых алгоритмов к характеристикам сигнала и помех.