2.7. Адаптивный обнаружитель сдвига гауссовского распределения с использованием классифицированных обучающих выборок

В задаче обнаружения сдвига гауссовского распределения: $H_0 : y_k \sim N(m_0, \sigma_0^2)$, $H_1^{(1)} : y_k \sim N(m_1, \sigma_0^2)$, $k = 1, ..., N$, по независимой выборке $y = (y_1, y_2, ..., y_N)$ рассмотрим случай обнаружения неизвестного сдвига Δ при неизвестных значениях m_0 и σ_0^2.

Для преодоления априорной неопределенности можно использовать адаптацию, которая возможна на основе обучения по *классифицированным* выборкам [12].

Пусть наряду с контрольной (анализируемой) выборкой $y = (y_1, y_2, ..., y_N)$ имеются две классифицированные обучающие выборки. Одна из них помеховая выборка $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, ..., x_{N_0}^{(0)})$ размера N_0, соответствующая гипотезе H_0, т. е. $x_i^{(0)} \sim N(m_0, \sigma_0^2)$. Вторая выборка $x^{(1)} = (x_1^{(1)}, x_2^{(1)}, ..., x_{N_1}^{(1)})$ размера N_1, соответствующая альтернативе $H_1^{(1)}$, т. е. $x_i^{(1)} \sim N(m_1, \sigma_0^2)$, представляет смесь помехи и полезного сигнала.

За основу берется байесовский алгоритм обнаружения, отвергающий гипотезу H_0, если логарифм отношения правдоподобия $l = \ln\Lambda$ превышает порог $\ln\Lambda_B$. В данной задаче $l = N\bar{y}(m_1 - m_0)/\sigma_0^2 - N(m_1^2 - m_0^2)/2\sigma_0^2$, где $\bar{y} = \sum_{k=1}^{N} y_k / N$ – среднее арифметическое контрольной выборки. Полагаем, что значение байесовского порога Λ_B известно.

Адаптивный алгоритм с обучением формирует оценки $\hat{m}_0$, $\hat{m}_1$ и $\hat{\sigma}_0^2$, которые подставляются вместо неизвестных значений при обработке контрольной выборки. Он принимает вид

$$\hat{l} = N\bar{y}(\hat{m}_1 - \hat{m}_0)/\hat{\sigma}_0^2 - N(\hat{m}_1^2 - \hat{m}_0^2)/2\hat{\sigma}_0^2 \geq \ln\Lambda_B,$$

где оценки неизвестных параметров формируются в блоке адаптации и вычисляются по формулам

$$\hat{m}_0 = \sum_{k=1}^{N_0} x_k^{(0)} / N_0 \,;\; \hat{m}_1 = \sum_{k=1}^{N_0} x_k^{(1)} / N_1 \,;$$

$$\hat{\sigma}_0^2 = [\sum_{k=1}^{N_0} (x_k^{(0)} - \hat{m}_0)^2 + \sum_{k=1}^{N_1} (x_k^{(1)} - \hat{m}_1)^2]/(N_0 + N_1 - 1).$$

Эти оценки являются несмещенными эффективными оценками указанных неизвестных параметров. Представленный алгоритм реализует *двухвыборочный метод* обнаружения, при этом обучающая выборка $x = (x^{(0)}, x^{(1)})$ включает $x^{(0)}$ и $x^{(1)}$.

Нетрудно видеть, что для критерия максимального правдоподобия $\ln\Lambda_{ML} = 0$, параметр σ_0^2 становится несущественным и не требует оценивания. Рассмотрим вначале этот случай.

Оценки $\hat{m}_0$ и $\hat{m}_1$ взаимно независимы и имеют гауссовские распределения с математическими ожиданиями m_0 и m_1, и дисперсиями σ_0^2/N_0 и σ_0^2/N_1.

Логарифм отношения правдоподобия можно записать в виде произведения [12]: $l = (N/2)z_1 z_2$, где $z_1 = (m_1 - m_0)/\sigma_0$, $z_2 = (2\bar{y} - m_1 - m_0)/\sigma_0$. Оценка логарифма отношения правдоподобия представляется в виде $\hat{l} = (N/2)\hat{z}_1\hat{z}_2 \geq \ln\Lambda_B$, где $\hat{z}_1 = (\hat{m}_1 - \hat{m}_0)/\sigma_0$ и $\hat{z}_2 = (2\bar{y} - \hat{m}_1 - \hat{m}_0)/\sigma_0$ включают оценки $\hat{m}_0$ и $\hat{m}_1$. Таким образом, адаптивный алгоритм реализуется двухканальной структурой корреляционного типа, которая показана на рис. 2.41. Для простоты полагалось $\sigma_0 = 1$. Структура чувствительна к совпадению знаков $\hat{m}_1 - \hat{m}_0$ и $\bar{y} - (\hat{m}_1 + \hat{m}_0)/2$, поэтому знак сдвига не является уже существенным параметром.

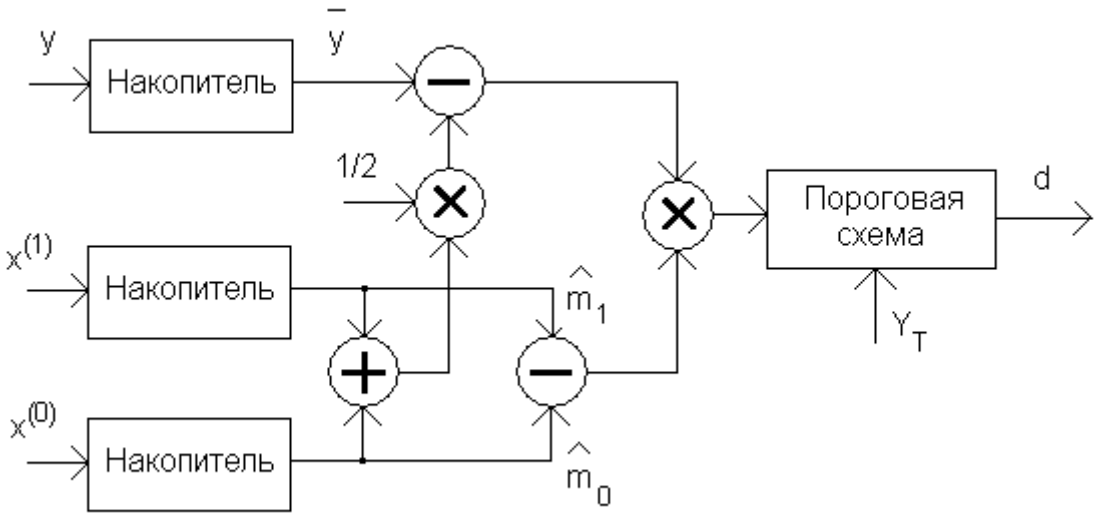

Рис. 2.41. Структура адаптивного обнаружителя

Величины $\hat{z}_1$ и $\hat{z}_2$ при фиксированном значении σ_0^2 являются совместно гауссовскими зависимыми случайными величинами. Частные (или *маргинальные*, от англ. marginal – краевой) распределения этих величин будут гауссовскими: $w_1(\hat{z}_1) \sim N(a_1, \sigma_1^2)$, $w_2(\hat{z}_2) \sim N(a_2, \sigma_2^2)$, где $a_1 = d = (m_1 - m_0)/\sigma_0$, $\sigma_1^2 = 1/N_0 + 1/N_1$.

Значение a_2 зависит от гипотезы: $a_2 = -d$ при гипотезе H_0, $a_2 = d$ при гипотезе H_1. Дисперсия статистики $\hat{z}_2$ равна $\sigma_2^2 = 4/N + 1/N_0 + 1/N_1$.

Коэффициент корреляции между $\hat{z}_1$ и $\hat{z}_2$ равен $R = \kappa_{11}/\sigma_1\sigma_2$, где ковариация $\kappa_{11} = M\{(\hat{z}_1 - a_1)(\hat{z}_2 - a_2)\} = 1/N_0 - 1/N_1$. В данном случае $\kappa_{11} = 1/N_0 - 1/N_1$. Отсюда следует, что при одинаковом размере обучающих выборок $(N_0 = N_1)$ случайные величины $\hat{z}_1$ и $\hat{z}_2$ некоррелированы, а в силу гауссовости, – и независимы. При неодинаковом размере выборок коэффициент корреляции будет невелик, если число выборок достаточно велико.

Плотность вероятности произведения $v = \hat{z}_1 \cdot \hat{z}_2$ выражается формулой [6]

$$w(v) = \int_{-\infty}^{\infty} w_1(u) w_2(v/u) \mathrm{d}u / |u|;$$

где $w_1(\cdot)$ и $w_2(\cdot)$ есть плотности вероятности величин $\hat{z}_1$ и $\hat{z}_2$.

Вероятность ложной тревоги равна вероятности превышения случайной величиной v нулевого порога, или

$$F = (1/\sqrt{2\pi}\sigma_2)\int_{-\infty}^{\infty} \exp\{-(u+d)^2/2\sigma_2^2\}\mathrm{d}u \cdot$$

$$\cdot \int_0^{\infty} (1/(\sqrt{2\pi}\sigma_1 |u|)) \exp\{-(v - du)^2/2\sigma_1^2 u^2\}\mathrm{d}v$$

Внутренний интеграл выражается через интеграл вероятности

$$\Phi(z) = (1/\sqrt{2\pi})\int_{-\infty}^{z} \exp(-t^2/2)\mathrm{d}t.$$

В результате получаем

$$F = (1/\sqrt{2\pi}\sigma_2)\int_{-\infty}^{\infty} \Phi(-du/\sigma_1 |u|) \exp\{-(u+d)^2/2\sigma_2^2\}\mathrm{d}u.$$

Разбивая пределы интегрирования на два интервала (от $-\infty$ до 0 и от 0 до ∞), получаем

$$F = \Phi(d/\sigma_1)\Phi(-d/\sigma_2) + \Phi(-d/\sigma_1)\Phi(d/\sigma_2).$$

В данном случае вероятность ложной тревоги равна вероятности пропуска, и эти вероятности совпадают с величиной суммарной вероятности ошибки (или среднего риска).

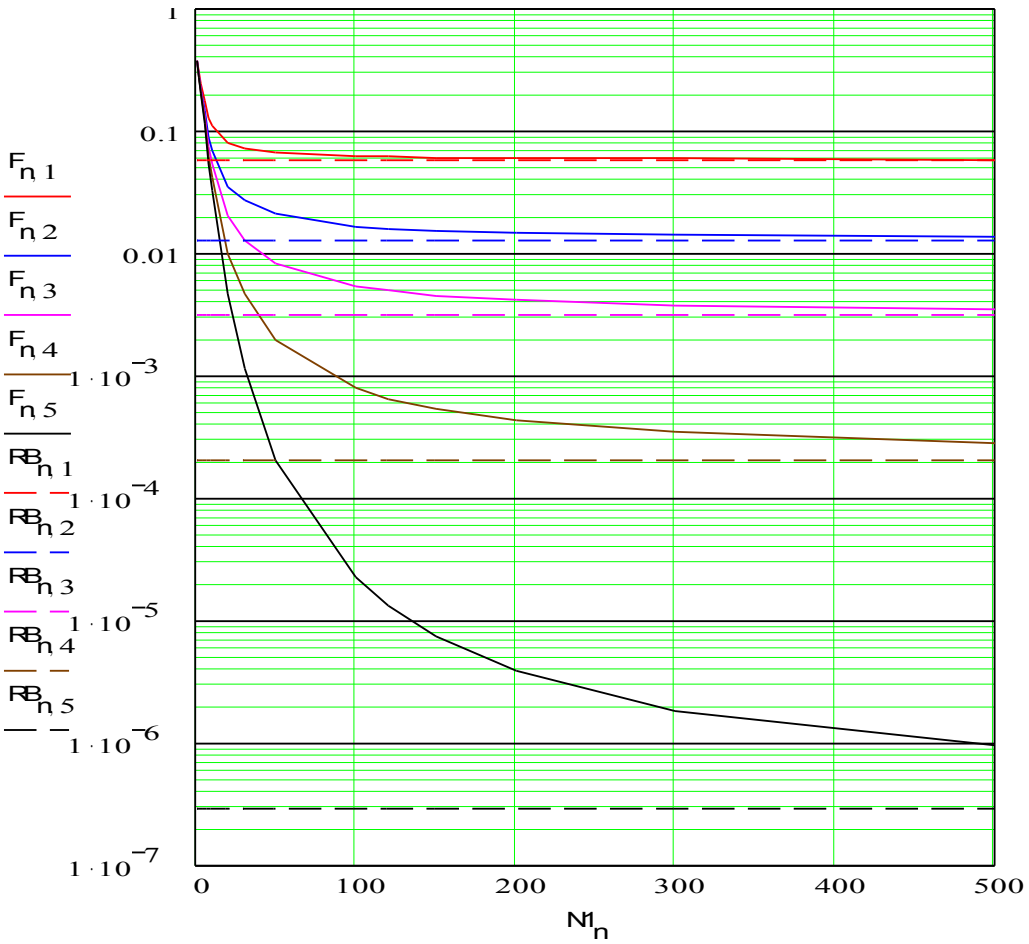

Рис. 2.42. Зависимость вероятности ошибки от размера обучающих выборок при $d = 1$, $N = 10, 20, 30, 50, 100$
nBA_Adapt_11.mcd

Зависимость суммарной вероятности ошибки p_Σ от размера обучающих выборок $(N_0 = N_1)$ показана на рис. 2.42 для $d = 1$ и разных размеров контрольной выборки $N = 10, 20, 30, 50, 100$ (кривые 1 –5 соответственно).

Графики на рис. 2.42 показывают, что при увеличении размера обучающих выборок вероятность ошибки стремится к значению, достигаемому в условиях известных параметров для данного размера контрольной выборки (которые показаны штриховыми линиями). Однако с увеличением размера контрольной выборки предельные вероятности ошибок уменьшаются, и для их достижения требуется все больше обучающих выборок.

Для достижения предельной суммарной вероятности ошибок при $N = 10$ требуется от 50 до 100 обучающих выборок. При дальнейшем их увеличении влияние обучения на вероятность ошибок становится менее ощутимым [12].

Если с увеличением размера N контрольной выборки значение d падает, так что обеспечивается постоянство дефлекции $d_N = \sqrt{N}d$, то в процессе обучения вероятности ошибок стремятся к одному и тому же значению: например, для $d_N = 4{,}65$ это значений равно 10^{-2} (рисунок 2.43).

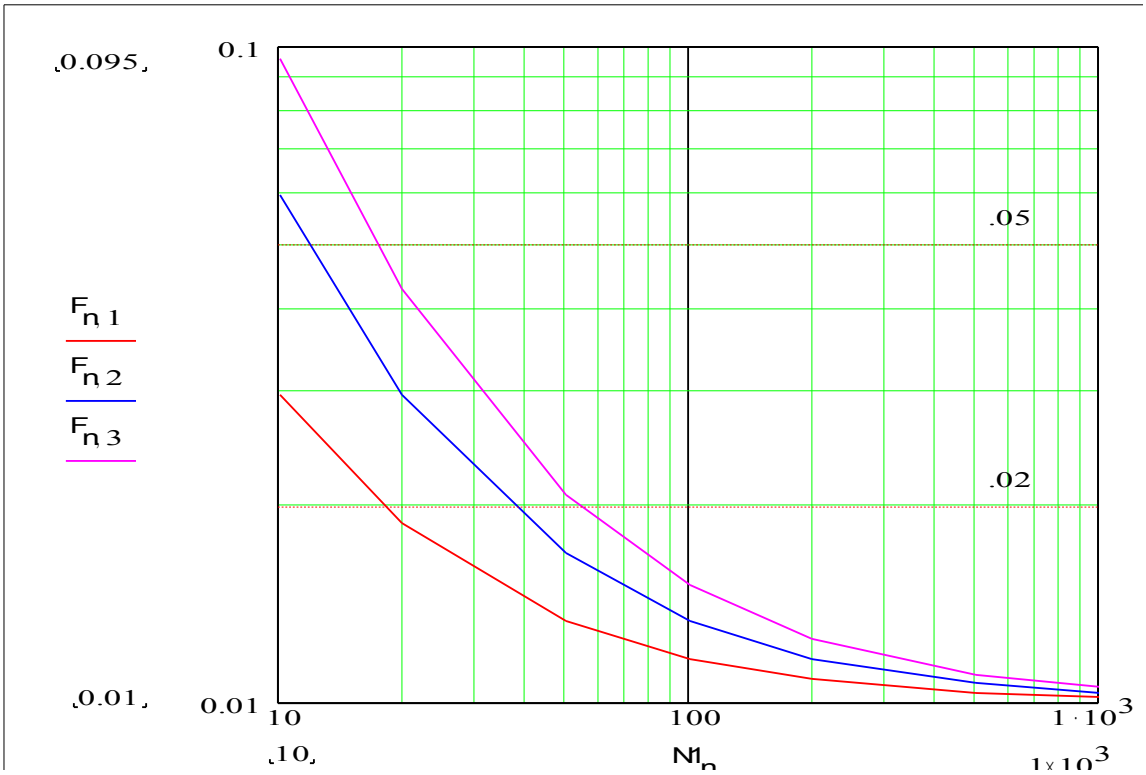

Рис. 2.43. Зависимость вероятности ошибки от размера обучающих выборок при $d_N = 4{,}65$, $N = 10, 20, 30$
nBA_Adapt_33.mcd

nBA_Adapt_33.mcd

Увеличение числа контрольных выборок при постоянной дефлекции d_N означает уменьшение значений величины d, которую можно рассматривать, как отношение сигнал/шум на одну выборку. Для $N = 10, 20, 30$ значения d оказываются равными 1,47, 1,04 и 0,849 соответственно. Уменьшение d резко снижает качество обучения. Поэтому кривые для больших N на рис. 2.43 соответствуют бо′льшим ошибкам, чем в предыдущем случае, когда значение d постоянно.

Таким образом, при большом размере контрольной выборки размер обучающих должен быть не меньшим, иначе процесс адаптации будет сопровождаться бо′льшими ошибками и снизит возможности основного алгоритма.

При малых обучающих выборках по сравнению с контрольной вероятности ошибок резко возрастают, и встает вопрос о целесообразности использования адаптации. Все будет зависеть от степени изменения параметров. При небольших изменениях не исключено, что неадаптивная структура даст меньшие вероятности ошибок, чем адаптивная ([1], подразд. 1.7).

Заметим, что если *заранее известен знак сдвига*, адаптивную обработку можно упростить. Например, при $m_1 > m_0$ получаем адаптивный вариант байесовского алгоритма максимального правдоподобия в виде $\bar{y} \geq (\hat{m}_1 + \hat{m}_0)/2$.

Рассмотрим построение адаптивного обнаружителя в случае неизвестных и разных (при гипотезе и при альтернативе) как математических ожиданий, так и дисперсий: $H_0 : y_k \sim N(m_0, \sigma_0^2)$, $H_1 : y_k \sim N(m_1, \sigma_1^2)$.

Адаптивный алгоритм с обучением по двум выборкам вытекает из общего алгоритма обнаружения в случае независимых наблюдений ([1], подразд. 1.5). Логарифм отношения правдоподобия (при $\sigma_1^2 \neq \sigma_0^2$) записывается в виде [12]:

$$l = \left(\frac{\sigma_1^2 - \sigma_0^2}{2\sigma_1^2 \sigma_0^2} \right) \sum_{k=1}^{N} (y_k - \rho)^2 - \left(\frac{N}{2} \right) \left(\ln \frac{\sigma_1^2}{\sigma_0^2} + \frac{(m_1 - m_0)^2}{(\sigma_1^2 - \sigma_0^2)} \right),$$

где $\rho = (m_0^2 \sigma_1^2 - m_1^2 \sigma_0^2)/(\sigma_1^2 - \sigma_0^2)$.

Неизвестные параметры, входящие в эти формулы, заменяются их оценками $\hat{m}_i = \sum_{k=1}^{N_i} x_k^{(i)} / N_i$, $\hat{\sigma}_i^2 = \sum_{k=1}^{N_i} (x_k^{(i)} - \hat{m}_i)^2 / (N_i - 1)$, $i = 0,1$, при этом получается оценка $\hat{l}$. Для случая равенства дисперсий предусматривается отдельный канал.

Адаптивный алгоритм принимает решение в пользу гипотезы H_1, если $\hat{l} \geq \ln \Lambda_B$. Его характеристики исследованы в [12] методом статистического моделирования.

Алгоритм предъявляет более жесткие требования к размерам обучающих выборок для достижения приемлемых вероятностей ошибок.

Если обучающая выборка $x = (x_1, x_2, ..., x_N)$ не классифицирована, то возможен вариант *самообучения*. С учетом случайности событий, связанных с появлением сигнала, наблюдения принадлежат смеси распределений $f(x) = p_0 \cdot f_0(x) + p_1 \cdot f_1(x)$ с математическим ожиданием $m = p_0 \cdot m_0 + p_1 \cdot m_1$. Если события равновероятны, то выборочное среднее $\bar{x}$ дает оценку для байесовского порога максимального правдоподобия $(m_0 + m_1)/2$.

Если известно, что $m_1 > m_0$, адаптивный алгоритм сводится к сравнению выборочных средних значений $\bar{y} \geq \bar{x}$ или $\bar{y} - \bar{x} \geq 0$. В данном случае алгоритм обнаружения оказывается инвариантным к одинаковым изменениям параметров сдвига и масштаба в обеих выборках.

Легко видеть, что для сравнения средних двух выборок можно вычислять *контраст* (отношение средних) и обнаруживать его возрастание, т. е. применять алгоритм $\bar{y}/\bar{x} \geq 1$. Однако данный вариант включает нелинейную операцию деления при формировании решающей статистики. Кроме того, отношение средних инвариантно только к изменениям параметра масштаба, и зависит от одинакового сдвига в обеих выборках.