O. V. Denisov, Statistical estimation of the significant arguments set of the binary vector-function with corrupted values,
Mat. Vopr. Kriptogr., 2014, Volume 5, Issue 4, 41–61

https://www.mathnet.ru/eng/mvk134

МАТЕМАТИЧЕСКИЕ ВОПРОСЫ КРИПТОГРАФИИ
2014 Т. 5 № 4 С. 41–61

УДК 519.719.2+519.233.2

Статистическая оценка множества существенных аргументов двоичной вектор-функции с искаженными значениями

О. В. Денисов

ООО «Центр сертификационных исследований», Москва

Получено 22.IV.2013

Пусть Θ — множество номеров существенных аргументов неизвестной двоичной вектор-функции со случайными равномерно распределенным аргументами и искаженными значениями. Предлагается алгоритм построения оценки Θ^* для Θ на основе статистических оценок спектра функции. Для некоторых классов функций (в частности, для векторных бент-функций и биективных отображений) получены асимптотические границы объема данных, при котором алгоритм начинает работать успешно, то есть $\mathbf{P}\{\Theta^* = \Theta\} \to 1$.

Ключевые слова: двоичная вектор-функция, существенные аргументы, оценки спектра функции.

Statistical estimation of the significant arguments set of the binary vector-function with corrupted values

O. V. Denisov

LLC "Certification Research Center", Moscow

Abstract. Let Θ be the set of significant arguments of the unknown binary vector-function with the random uniformly distributed arguments and corrupted values. Algorithm for constructing the estimate Θ^* of Θ based on statistical estimates of function spectrum is proposed. For some function classes (particularly, for vectorial bent-functions and bijective mappings) we get asymptotic bounds of the data size sufficient for the successful work of the algorithm, i.e. $\mathbf{P}\{\Theta^* = \Theta\} \to 1$.

Keywords: binary vector-function, essential arguments, function spectrum estimations.

Citation: *Mathematical Aspects of Cryptography*, 2014, vol. 5, no. 4, pp. 41–61 (Russian)

1. Постановка задачи, обозначения и результаты

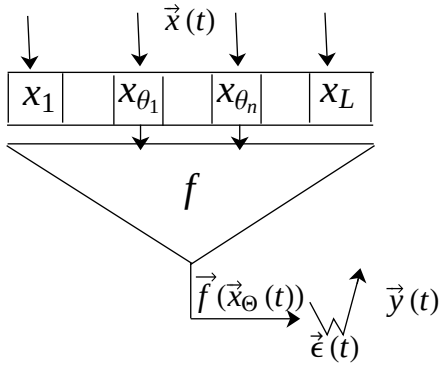

Рис. 1. Схема двоичного преобразования с искажением

В некоторых задачах криптографии и теории информации возникает схема, представленная на рис. 1, где для наблюдения доступны вход $\vec{x}(t)$ и выход $\vec{y}(t)$ схемы, $1 \le t \le N$. Требуется определить множество $\Theta = \{\theta_1, \ldots, \theta_n\}$ номеров существенных входов.

В данной статье предлагается метод решения этой задачи в случае, когда входные векторы (аргументы) являются случайными независимыми и равномерно распределенными векторами:

$$\vec{x}(t) = (x_1(t), \ldots, x_L(t)) \sim U(\mathbb{Z}_2^L).$$

Здесь и далее $(\mathbb{Z}_2, \oplus)$ — группа вычетов по модулю 2, $\mathbb{Z}_2^L$ — множество двоичных векторов размерности L (L-я декартова степень множества $\mathbb{Z}_2 = \{0, 1\}$). Преобразующая вектор-функция $\vec{f} : \mathbb{Z}_2^n \to \mathbb{Z}_2^m$ считается неизвестной, но известно число ее аргументов $n = |\Theta| < L$. Векторы искажений

$$\vec{\varepsilon}(1), \ldots, \vec{\varepsilon}(N) \sim \vec{\varepsilon} \nsim U(\mathbb{Z}_2^m)$$

одинаково распределены и не зависят друг от друга, а также от всех векторов аргументов. Распределение $\vec{\varepsilon}$ неравномерно и неизвестно. Искажения аддитивно и покоординатно накладываются на значения вектор-функции, и на выходе наблюдаются случайные векторы

$$\vec{y}(t) = \vec{f}(\vec{x}_\Theta(t)) \oplus \vec{\varepsilon}(t), \quad 1 \le t \le N.$$

Здесь и далее для произвольного вектора $\vec{x} = (x_1, \ldots, x_M)$ и непустых $I = \{i_1, \ldots, i_r\} \subset \{1, \ldots, M\}$ через

$$\vec{x}_I = (x_{i_1}, \ldots, x_{i_r})$$

обозначаем подвектор вектора $\vec{x}$, состоящий из координат с номерами из I.

Условия, перечисленные в последнем абзаце, далее называем условиями **(ВФ)**. Они задают статистическую параметрическую модель, в которой неизвестны параметры Θ, $\vec{f}$ и распределение $\vec{\varepsilon}$. Далее будет предложен алгоритм построения статистической оценки Θ^* для Θ путем поэтапного вычисления последовательности множеств

$$\Theta^*(0) = \varnothing \subset \Theta^*(1) \subset \ldots \subset \Theta^*(\tau) = \Theta^*$$

со случайным числом этапов τ. На идейном уровне работу алгоритма можно пояснить так. На этапе r для всех $i \notin \Theta^*(r-1)$ вычисляются случайные векторы $\vec{S}_N(i, \Theta^*(r-1))$ размерности $2^{|\Theta^*(r-1)|}(2^m - 1)$, координаты которых являются (с точностью до линейного преобразования) значением выборочной ковариации случайной величины x_i и линейной комбинации значений входов с номерами из $\Theta^*(r-1)$ и выходов. Если евклидова норма такого вектора не меньше значения некоторой граничной функции, то принимается решение о включении i-го входа в $\Theta^*(r)$.

В разделе 2 доказывается (теорема 1), что при надлежащем выборе граничной функции с вероятностью, не меньшей заданной величины, оценка Θ^* для любого N не содержит номеров из дополнения

$$\Theta^c = \{1, \ldots, L\} \setminus \Theta,$$

и временная сложность построения Θ^* не превосходит величины порядка $2^{n+m}L$.

Далее в разделах 3 и 4 исследуется зависимость вероятности успешной ($\Theta^* = \Theta$) работы метода от объема материала (объема данных) N. В основной теореме 2 получены асимптотические оценки объема материала, при котором на первом же этапе алгоритм правильно выделяет все существенные переменные, либо, напротив, не выделяет ни одной переменной как существенной.

При поиске порога объема (threshold), то есть значения $N = N_{\text{th}}$, при котором метод «начинает» работать успешно, выбран следующий асимптотический подход, классический в вероятностной комбинаторике. Рассматривается схема серий, в которой параметры n, m, $\vec{f}$, L, N, а также распределение $\vec{\varepsilon}$ зависят от номера серии. Функция N_{th} называется *порогом объема материала* для события $A = \{\Theta^* = \Theta\}$, если

$$\lim \mathbf{P}(A) = \begin{cases} 1, & \text{если } \underline{\lim}\, N/N_{\text{th}} > 1, \\ 0, & \text{если } \overline{\lim}\, N/N_{\text{th}} < 1. \end{cases} \tag{1}$$

Заметим, что такая функция определена неоднозначно; любая эквивалентная ей функция $N^*_{\text{th}} \sim N_{\text{th}}$ (то есть такая что $\lim N_{\text{th}}/N_{\text{th}} = 1$) также является порогом.

Вероятность успешной работы существенно зависит от спектра функции $\vec{f}$ и спектра распределения шума $\vec{\varepsilon}$. Строго эти понятия определяются ниже в разделе 1.1. Порог объема материала найден в явном виде в следствии 2.1, условия которого выполнены, в частности, в случае шума с однородным спектром и спектрально 1-однородных функций (определения даны ниже в разделе 4).

Определению спектрально 1-однородных функций удовлетворяют, в частности, бент-функции и биективные отображения, а в булевом случае ($m = 1$) — симметрические пороговые функции, в том числе мажоритарные. Для них в разделе 4 найдены явные выражения для порогов объема материала.

Приведём необходимые определения и факты.

Спектры распределений и функций

Далее для вектора $\vec{x} = (x_1, \ldots, x_M) \in \mathbb{R}^M$ через

$$\|\vec{x}\| = x_1 + \ldots + x_M, \quad |\vec{x}| = (x_1^2 + \ldots + x_M^2)^{1/2}$$

обозначаем соответственно *вес вектора* и его евклидову норму; для $\vec{x} \in \mathbb{Z}_2^M$ вес обозначает число ненулевых координат вектора. Для множества G через $|G|$ обозначается его мощность.

Символы $\|$ будут применяться также при описании матриц: равенство

$$A = \|a_{i,j}\|_{i \in X, j \in Y}$$

задает матрицу, элементы которой индексируются значениями из множеств X, Y. Строки и столбцы матрицы A обозначаем $\vec{A}_i$ и $A^{\downarrow}_j$.

Введем спектральные характеристики распределений двоичных векторов и отметим некоторые их известные свойства. Сначала сделаем это для распределения шума, а затем для совместного распределения входов и выходов.

Случайная величина η называется *индикаторной*, если она принимает только значения 0 и 1. *Преобладанием нуля* в ее распределении называется величина

$$\mathbf{d}(\eta) = \mathbf{E}(-1)^\eta = \mathbf{P}\{\eta = 0\} - \mathbf{P}\{\eta = 1\} = 2\mathbf{P}\{\eta = 0\} - 1 \in [-1, 1].$$

Для случайного вектора $\vec{\xi} = (\xi_1, \ldots, \xi_M)$ со значениями в $\mathbb{Z}_2^M$ через

$$\phi_{\vec{\xi}}(J) = \mathbf{E}(-1)^{\|\vec{\xi}_J\|} = \mathbf{E}(-1)^{\xi_{j_1}+\ldots+\xi_{j_l}}, \quad J = \{j_1, \ldots, j_l\} \subset \{1, \ldots, M\},$$

обозначим значения *характеристической функции* [1] (х.ф.) распределения $\vec{\xi}$. Эти значения по модулю не превосходят 1 и равны преобладаниям нуля $\mathbf{d}(\xi_{j_1} \oplus \ldots \oplus \xi_{j_l})$ в распределениях линейных комбинаций. Набор значений $\phi_{\vec{\xi}}(J)$ по всем J называется *спектром распределения* $\vec{\xi}$.

Х.ф. распределений двоичных векторов сохраняют многие свойства обычных х.ф. (см. [1]), в том числе мультипликативное свойство: если случайные векторы $\vec{\xi}$ и $\vec{\eta}$ размерности M независимы, то

$$\phi_{\vec{\xi}\oplus\vec{\eta}}(J) = \phi_{\vec{\xi}}(J)\phi_{\vec{\eta}}(J), \ J \subset \{1, \ldots, M\}.$$

Кроме того, справедливы специфические *спектральный критерий равномерности распределения*:

$$\vec{\xi} \sim U(\mathbb{Z}_2^M) \Longleftrightarrow \phi_{\vec{\xi}}(J) = 0 \text{ для всех } \varnothing \neq J \subset \{1, \ldots, M\},$$

(из которого следует, что для шума $\vec{\varepsilon}$ при наших условиях $\phi_{\vec{\varepsilon}}(J) \neq 0$ хотя бы для одного непустого J), а также *равенство Парсеваля* [1]: для любой булевой функции $g : \mathbb{Z}_2^M \to \mathbb{Z}_2$ при случайном равномерно распределенном аргументе $\vec{x}$ для случайного вектора $(\vec{x}, g(\vec{x}))$ размерности $M+1$ имеем

$$\sum_{J \subset \{1,\ldots,M\}} \phi^2_{(\vec{x},g(\vec{x}))}(J \cup \{M+1\}) = 1.$$

Будем далее обозначать для вектор-функции

$$\vec{f} = \big(f_1(\vec{x}), \ldots, f_m(\vec{x})\big)$$

через

$$\vec{xf} = \big(x_1, \ldots, x_n, f_1(\vec{x}), \ldots, f_m(\vec{x})\big)$$

случайный вектор размерности $n+m$, составленный из аргументов и значений функции, при равномерном распределении аргумента $\vec{x} \sim U(\mathbb{Z}_2^n)$.

Теперь рассмотрим значения х.ф. вектора $\vec{xf}$:

$$\phi_{\vec{xf}}(I, J) = \mathbf{E}(-1)^{\|\vec{x}_I\|+\|\vec{f}_J(\vec{x})\|} = 2^{-n} \sum_{\vec{a}\in\mathbb{Z}_2^n} (-1)^{\|\vec{a}_I\|+\|\vec{f}(\vec{a})_J\|} =$$

$$= 2^{-n} \sum_{\vec{a}\in\mathbb{Z}_2^n} (-1)^{a_{i_1}+\ldots+a_{i_r}+f_{j_1}(\vec{a})+\ldots+f_{j_l}(\vec{a})},$$

$$I = \{i_1, \ldots, i_r\} \subset \{1, \ldots, n\}, \quad J = \{j_1, \ldots, j_l\} \subset \{1, \ldots, m\}.$$

Эти величины являются коэффициентами в разложениях Фурье псевдобулевых функций $(-1)^{\|\vec{f}_J(\vec{x})\|}$, и их называют *спектром вектор-функции* $\vec{f}$. С другой стороны, значения х.ф. являются нормированными коэффициентами Уолша-Адамара:

$$\phi_{\vec{xf}}(I,J) = 2^{-n} W_f(\vec{u},\vec{v}), \quad W_f(\vec{u},\vec{v}) = \sum_{\vec{a}\in\mathbb{Z}_2^n} (-1)^{\langle\vec{u},\vec{a}\rangle+\langle\vec{v},\vec{f}(\vec{a})\rangle}, \tag{2}$$

где $\langle\vec{x},\vec{y}\rangle = x_1y_1 + \ldots + x_ny_n$ — скалярное произведение, если множества единичных координат векторов $\vec{u}\in\mathbb{Z}_2^n$, $\vec{v}\in\mathbb{Z}_2^m$ суть I и J соответственно.

Введем *спектральную матрицу функции* $\vec{f}$ размера $2^n \times (2^m-1)$, состоящую из значений х.ф.:

$$\Phi = \Phi(\vec{f}) = \|\phi_{\vec{xf}}(I,J)\|_{I\subset\{1,\ldots,n\},\varnothing\neq J\subset\{1,\ldots,m\}}.$$

Согласно равенству Парсеваля, евклидовы нормы всех столбцов равны 1. Столбец, соответствующий индексу $J=\varnothing$, равен нулю при $\vec{x}\sim U(\mathbb{Z}_2^n)$ и поэтому здесь не участвует.

Введем также *внешнюю спектральную матрицу*, состоящую из значений х.ф. совместного распределения $\vec{x},\vec{y}$ внешних сигналов схемы:

$$\Psi = \|\phi_{\vec{x},\vec{y}}(I,J)\|_{I\subset\{1,\ldots,L\},\varnothing\neq J\subset\{1,\ldots,m\}}, \quad \phi_{\vec{x},\vec{y}}(I,J) = \mathbf{E}(-1)^{\|\vec{x}_I\|+\|\vec{y}_J\|},$$

размера $2^L\times(2^m-1)$. Из ниже приведенной формулы (3) и равенства Парсеваля следует, что

$$|\Psi_J^{\downarrow}| = |\phi_{\vec{\varepsilon}}(J)| \le 1, \quad \varnothing\neq J\subset\{1,\ldots,m\},$$

то есть нормы столбцов матрицы Ψ уже не обязательно равны 1.

Из условий **ВФ** (равномерность распределения аргументов, независимость их от шума) и мультипликативного свойства х.ф. вытекает, что для всех $I\subset\{1,\ldots,L\}$, $J\subset\{1,\ldots,m\}$

$$\phi_{\vec{x},\vec{y}}(I,J) = \begin{cases} 0 \text{ при } I\cap\Theta^c\neq\varnothing, \\ \phi_{\vec{xf}}(\{l_1,\ldots,l_r\},J)\phi_{\vec{\varepsilon}}(J) \text{ при } I=\{\theta_{l_1},\ldots,\theta_{l_r}\}\subset\Theta. \end{cases} \tag{3}$$

Таким образом, если I содержит хотя бы один элемент множества Θ^c номеров несущественных входов, то соответствующая строка равна нулю: $\vec{\Psi}_I = \vec{0}$. Если же I состоит только из номеров существенных входов,

то $\vec{\Psi}_I$ получена умножением элементов соответствующей строки матрицы Φ на значения х.ф. шума, причем значения х.ф. не все нулевые. Поэтому если статистические оценки показывают, что норма $\vec{\Psi}_I$ ненулевая, то $I \subset \Theta$. На этом основывается предлагаемый далее алгоритм.

При статистических оценках будем использовать спектральные статистики (суммы несмещенных оценок элементов матрицы Ψ)

$$S_N(I,J) = \sum_{1\le t\le N} s(t,I,J), \quad s(t,I,J) = (-1)^{\|\vec{x}_I(t)\|+\|\vec{y}_J(t)\|},$$
$$\mathbf{E}s(t,I,J) = \phi_{x,y}(I,J),\ \mathbf{D}s(t,I,J) = 1-\phi^2_{x,y}(I,J). \tag{4}$$

Из статистик $s(t,I,J)$ и $S_N(I,J)$ составим матричные статистики: ± 1-матрицы

$$\vec{s}(t,i,G) = \|s(t,\{i\}\cup I,J)\|_{I\subset G,\varnothing\ne J\subset\{1,\dots,m\}}$$

состоящие из

$$M(G) = 2^{|G|}(2^m-1)$$

элементов, и их суммы

$$\vec{S}_N(i,G) = \|S_N(\{i\}\cup I,J)\|_{I\subset G,\varnothing\ne J\subset\{1,\dots,m\}} = \sum_{1\le t\le N}\vec{s}(t,i,G). \tag{5}$$

Мы используем обозначения со знаком вектора, так как при оценках распределения нам удобнее интерпретировать эти случайные матрицы как случайные векторы размерности $M(G)$.

В качестве G будут рассматриваться множества номеров

$$G = \Theta^*(r) \subset \{1,\dots,L\}, r\ge 0,$$

выделенных как элементы Θ^* при завершении r-го этапа алгоритма.

Если $i\notin\Theta$, то с учетом (3) имеем

$$\mathbf{E}\vec{s}(t,i,G) = \vec{0},\ |\vec{s}(t,i,G)| = \sqrt{M(G)},\ \mathbf{E}(|\vec{s}(t,i,G)|^2) = M(G). \tag{6}$$

Для оценок распределения $|\vec{S}_N(i,G)|$ будем применять следующий вариант [2] неравенства Ю.В.Прохорова [4] для «хвостов» распределения нормы сумм независимых центрированных случайных векторов. Пусть $S_N = \sum_{1\le i\le N} X_i$, где $X_1,\dots,X_N$ — независимые одинаково распределенные случайные векторы со значениями в евклидовом или гильбертовом пространстве, и $\mathbf{E}X_i = 0$, $|X_i|\le A$, $\sigma^2 = \mathbf{E}(|X_i|^2)$. Тогда

$$\mathbf{P}\left\{|S_N|\ge B\sqrt{N}\right\} \le 2\exp\left\{-\tfrac{B^2}{2\sigma^2}\left(1+\tfrac{AB}{3\sigma^2\sqrt{N}}\right)^{-1}\right\}, \quad B>0. \tag{7}$$

Из (7) с учетом (6) и вытекающего из (6) равенства $\sigma^2 = M(G)$ получаем, что для всех $i \notin \Theta$

$$\mathbf{P}\left\{|\vec{S}_N(i,G)| \geq B\sqrt{NM(G)}\right\} \leq 2\exp\left\{-\frac{B^2}{2}\left(1+\frac{B}{3\sqrt{N}}\right)^{-1}\right\}, \quad B>0. \tag{8}$$

2. Алгоритм построения Θ^*, оценки сложности и вероятности успешной работы

Предлагается следующий алгоритм построения статистической оценки Θ^* для множества Θ на основе спектральных векторов (5). Множества $Add(r)$ обозначают номера входов, включенных в Θ^* на r-м этапе алгоритма.

Алгоритм построения множества Θ^*

(*граничная функция (граница)* $B(G)$ — параметр алгоритма)

1. Инициализация: полагаем $\Theta^*(0) = \varnothing$, $r = 1$.

2. Этап номер r:

а) Полагаем $Add(r) = \varnothing$.

б) Для всех $i \notin \Theta^*(r-1)$ вычисляем $\vec{S}_N(i, \Theta^*(r-1))$ и добавляем i в $Add(r)$, если

$$\left|\vec{S}_N(i,\Theta^*(r-1))\right| \geq \sqrt{NM(\Theta^*(r-1))}B(\Theta^*(r-1)).$$

в) Полагаем $\Theta^*(r) = \Theta^*(r-1) \cup Add(r)$.

3. Проверка окончания: если $Add(r) = \varnothing$ или $|\Theta^*(r)| \geq n$, то полагаем $\tau = r$, $\Theta^* = \Theta^*(\tau)$ и завершаем алгоритм. В противном случае увеличиваем r на 1 и повторяем шаг 2.

Везде далее в алгоритме будем использовать граничную функцию

$$B_1(G) = \sqrt{(2+\gamma)\ln\left\{\tbinom{n}{|G|}2^{|G|+2}(L-n)/\alpha\right\}}, \tag{9}$$

$$\gamma^2 = \gamma^2(n,L,\alpha,N) = \tfrac{1}{N}\max_{0\leq k\leq n}\ln\left\{\tbinom{n}{k}2^{k+2}(L-n)/\alpha\right\}$$

для всех $0 < \alpha < 1$. Она выбрана с учетом (8) так, чтобы далее была выполнена оценка (11). Начальное значение границы обозначаем

$$B_1 = B_1(\varnothing) = \sqrt{(2+\gamma)\ln\{4(L-n)/\alpha\}}.$$

Так как условие $\binom{n}{k+1}2^{k+1} \geq \binom{n}{k}2^k$ равносильно условию $k \leq \frac{2n-1}{3}$, то

$$\gamma^2 = \tfrac{1}{N}\ln\left\{\tbinom{n}{k_0}2^{k_0}\tfrac{4(L-n)}{\alpha}\right\}, \quad k_0 = \left[\tfrac{2n-1}{3}\right] + 1 = \left[\tfrac{2(n+1)}{3}\right],$$

а функция $B_1(G)$ при $1 \leq |G| \leq (2n-1)/3$ возрастает с ростом мощности G.

Из неравенства $\binom{n}{k_0}2^{k_0} < \sum_{0\leq k\leq n}\binom{n}{k}2^k = 3^n$ получаем оценку $\gamma^2 < \frac{1}{N}\left(n\ln 3 + \ln\frac{4(L-n)}{\alpha}\right)$, и поэтому

$$\gamma \leq 1/4, \text{ если } N \geq 16\left(n\ln 3 + \ln\tfrac{4(L-n)}{\alpha}\right);$$
$$\gamma \to 0, \text{ если } \left(n + \ln\tfrac{L-n}{\alpha}\right)/N \to 0. \tag{10}$$

Далее будем рассматривать только случаи, когда γ мало́.

Временно́й сложностью C_{time} алгоритма будем считать случайную величину, равную общему количеству вычисляемых статистик $S_N(I,J)$.

Заметим, что $\mathbf{E}\vec{S}_N(i,G) = N\mathbf{E}\vec{s}(t,i,G)$ согласно (5). Поэтому при $i \notin \Theta$ имеем $\mathbf{E}\vec{S}_N(i,G) = \vec{0}$; если же $i \in \Theta$ и $\mathbf{E}\vec{s}(t,i,G) \neq \vec{0}$, то математическое ожидание суммы растет пропорционально N, и можно ожидать, что при достаточно больших N вероятность события п.2б) алгоритма будет большой. Таким образом, выбор граничной функции регулирует характеристики статистики Θ^*: при увеличении $B(G)$ вероятность попадания в Θ^* ложных номеров уменьшается, но уменьшается также вероятность попадания истинных. Покажем, что при выборе границы вида (9) вероятность наличия в Θ^* номеров фиктивных переменных при любом N не будет превосходить α.

Теорема 1. *Пусть выполнены условия* **ВФ**, $0 < \alpha < 1$ *и* $\gamma \leq 1/4$. *Тогда при любой функции* $\vec{f}$ *справедлива оценка*

$$\mathbf{P}\left\{\Theta^* \subset \Theta,\ C_{time} \leq 2^n(2^m-1)(L-n+1)\right\} \geq 1-\alpha.$$

Доказательство. Если $\Theta^* \subset \Theta$, то все опробуемые множества I удовлетворяют условию $|I \setminus \Theta| \leq 1$, и тогда

$$C_{time} \leq 2^{|\Theta^*|}(2^m-1)\left(1 + \tbinom{L-n}{1}\right) \leq 2^n(2^m-1)(L-n+1).$$

Поэтому достаточно доказать, что вероятность события $\{\Theta^* \not\subseteq \Theta\}$ не превосходит α.

При условиях теоремы для любого G мощности k имеем

$$1 + \tfrac{B_1(G)}{3\sqrt{N}} \leq 1 + \tfrac{1}{3}\sqrt{(2+\tfrac{1}{4})\tfrac{1}{N}\ln\left\{\tbinom{n}{k}2^{k+2}(L-n)/\alpha\right\}} \leq 1 + \tfrac{\gamma}{2}.$$

Тогда для всех $i \notin \Theta$ согласно (8) выполнено

$$\mathbf{P}\left\{|\vec{S}_N(i,G)| \geq B_1(G)\sqrt{NM(G)}\right\} \leq$$
$$\leq 2\exp\left\{-\tfrac{2+\gamma}{2}\ln\left\{\tbinom{n}{k}2^{k+2}(L-n)/\alpha\right\}(1+\gamma/2)^{-1}\right\} =$$
$$= \frac{2\alpha}{\binom{n}{k}2^{k+2}(L-n)}. \quad (11)$$

Если $\Theta^* \not\subseteq \Theta$, то существуют такие $G = \Theta^*(r-1) \subset \Theta$ и $i \in \Theta^c$, при которых норма вектора $\vec{S}_N(i,G)$ впервые превысила порог. Тогда с учетом последней оценки имеем

$$\mathbf{P}\{\Theta^* \not\subseteq \Theta\} \leq \sum_{0\leq k\leq n}\sum_{G\subset\Theta,|G|=k}\sum_{i\in\Theta^c}\frac{\alpha}{\binom{n}{k}2^{k+1}(L-n)} < \alpha\sum_{k\geq 0}2^{-k-1} = \alpha.$$

□

3. Асимптотические оценки объема материала при одноэтапной работе

Получим оценки объема материала, при котором на первом же этапе алгоритм правильно выделяет все существенные переменные, либо, напротив, не выделяет ни одной переменной как существенной. Заметим, что в обоих случаях работа завершается на первом этапе. Полученные оценки будут определяться значениями х.ф. $\phi_{\vec{xf}}(I,J)$ при $|I|=1$, а также спектральными свойствами шума.

Обозначим для краткости при $i \notin G$

$$\vec{\mu}(i,G) = \mathbf{E}\vec{s}(t,i,G).$$

Этот вектор размерности $M(G)$ может быть представлен конкатенацией строк $\vec{\Psi}_{\{i\}\cup I}$ по всем $I \subset G$, и

$$\mathbf{E}\vec{S}_N(i,G) = N\vec{\mu}(i,G).$$

При $G = \varnothing$ и $i = \theta_l$ согласно (4) и (3) имеем

$$|\vec{\mu}(\{\theta_l\},\varnothing)|^2 = \sum_{\varnothing\neq J\subset\{1,\ldots,m\}}\phi^2_{\vec{xf}}(\{\theta_l\},J)\phi^2_{\varepsilon}(J) = |\vec{\Psi}_{\{\theta_l\}}|^2. \quad (12)$$

Обозначим через

$$\psi_{\min} = \min_{1\leq l\leq n}|\vec{\mu}(\{\theta_l\},\varnothing)|, \quad \psi_{\max} = \max_{1\leq l\leq n}|\vec{\mu}(\{\theta_l\},\varnothing)|$$

минимальную и максимальную нормы строк внешней спектральной матрицы Ψ, отвечающих одноэлементным наборам существенных входов.

Далее рассматривается схема серий, в которой параметры $n, m,\ \vec{f},\ L, \alpha,\ N$, а также распределение $\vec{\varepsilon}$ зависят от номера серии. Считаем, что при неограниченном возрастании номера серии

$$n \to \infty,\ \alpha \to 0,\ \left(n + \ln \tfrac{L-n}{\alpha}\right)/N \to 0, \tag{13}$$

и исследуем предел вероятности успешной работы алгоритма. В силу (10) из (13) следует, что $\gamma \to 0$, и тогда

$$B_1 \sim \sqrt{2\ln(4(L-n)/\alpha)}. \tag{14}$$

Теорема 2. *Пусть при условиях* **ВФ** *в схеме серий выполнены условия* (13)*, и* $\psi_{\min} > 0$ *в каждой серии. Тогда:*
1. Если $B_1 > \sqrt{b \ln n}$ *для некоторой константы* $b > 2$ *и*

$$N \le N_1 = \frac{(2^m - 1)\left(B_1 - \sqrt{b \ln n}\right)^2}{\psi_{\max}^2},$$

то $\mathbf{P}\{\Theta^*(1) = \varnothing\} \to 1$.
2. Если для некоторой константы $b > 2$

$$N \ge N_2 = \frac{(2^m - 1)\left(B_1 + \sqrt{b \ln n}\right)^2}{\psi_{\min}^2},$$

то $\mathbf{P}\{\Theta^*(1) = \Theta\} \to 1$.

Доказательство. **1**. Из последнего условия в (13) с учетом (10) имеем $\gamma \to 0$, и для всех серий с достаточно большим номером выполнены условия теоремы 1. Тогда для любого $i \in \Theta^c$ согласно оценке (11) имеем

$$\mathbf{P}\left\{|\vec{S}_N(i,\varnothing)| \ge B_1\sqrt{NM(\varnothing)}\right\} \le \tfrac{2\alpha}{4(L-n)},$$
$$\mathbf{P}\left\{\Theta^*(1) \cap \Theta^c \ne \varnothing\right\} \le \sum_{i \in \Theta^c} \tfrac{\alpha}{2(L-n)} = \alpha/2 \to 0. \tag{15}$$

2. Для случая $i \in \Theta$ и произвольного G выведем аналогично (11) оценку вероятности уклонения $\vec{S}_N = \vec{S}_N(i,G)$ от математического ожидания. Здесь случайный вектор $\vec{S}_N - N\vec{\mu}$, $\vec{\mu} = \vec{\mu}(i,G)$, представляется суммой независимых одинаково распределенных случайных векторов $\vec{X}_t = \vec{s}(t,i,G) - \vec{\mu}$ с нулевым средним, ограничением нормы

$$|\vec{X}_t| \le |\vec{s}| + |\vec{\mu}| \le A = 2\sqrt{M(G)}$$

и средним квадратом нормы

$$\sigma^2 = \mathbf{E}|\vec{X}_t|^2 = \sum_{I\subset G,J} \mathbf{D}s(I,J) = M(G) - |\vec{\mu}|^2$$

согласно (4).

Тогда из (7) находим

$$\mathbf{P}\left\{|\vec{S}_N - N\vec{\mu}| \ge B\sqrt{NM(G)}\right\} \le 2\exp\left\{-\tfrac{B^2}{2}\left(1+\tfrac{2B}{3\sqrt{N}}\right)^{-1}\right\} \qquad (16)$$

при всех $B>0$. В частности, при $B=\sqrt{b\ln n}$ ниже потребуется оценка: равномерно относительно $i\in\Theta$ выполняется

$$\mathbf{P}\left\{|\vec{S}_N - N\vec{\mu}| \ge \sqrt{NM(G)b\ln n}\right\} = o(1/n). \qquad (17)$$

Ее можно обосновать так: рассмотрим функцию $\kappa = 1/\left(1+\frac{2\sqrt{b\ln n}}{3\sqrt{N}}\right)$. Из последнего условия в (13) следует, что $n/N\to 0$, $\kappa\to 1$. Поэтому для произвольной фиксированной константы $1<c<b/2$, начиная с некоторой серии, выполнено $\frac{b}{2}\kappa \ge c$, и левая часть (17) согласно (16) не превосходит величины

$$2\exp\left(-\tfrac{b\ln n}{2}\kappa\right) \le 2n^{-c} = o(n^{-1}).$$

3. Далее ограничимся случаем $i\in\Theta$, $G=\varnothing$, и обозначим $M = M(\varnothing) = 2^m-1$. Тогда

$$\psi_{\min} \le |\vec{\mu}| \le \psi_{\max}.$$

При условиях п.1 имеем

$$|N\vec{\mu}| \le \sqrt{N}\sqrt{N_1}\psi_{\max} = \sqrt{NM}(B_1 - \sqrt{b\ln n}),$$

то есть среднее значение статистики $\vec{S}_N/\sqrt{NM}$ лежит в круге, радиус которого существенно меньше границы критерия. Тогда неравенство $|\vec{S}_N| \ge B_1\sqrt{NM}$ влечет неравенства

$$|\vec{S}_N - N\vec{\mu}| \ge |\vec{S}_N| - |N\vec{\mu}| \ge \sqrt{NM}\sqrt{b\ln n},$$

и поэтому согласно (17) имеем $\mathbf{P}\left\{|\vec{S}_N| \ge B_1\sqrt{NM}\right\} = o(1/n)$ равномерно относительно $i\in\Theta$. Отсюда с учетом оценки (15) получаем

$$\mathbf{P}\{\Theta^*\ne\varnothing\} \le \sum_{1\le i\le L} \mathbf{P}\left\{|\vec{S}_N(i,\varnothing)| \ge B_1\sqrt{NM}\right\} \le$$

$$\le \alpha/2 + \sum_{i\in\Theta} o(1/n) = o(1).$$

Утверждение первого пункта теоремы доказано.

4. При условиях п.2 при $i \in \Theta$ имеем аналогично

$$|N\vec{\mu}| \geq \sqrt{N}\sqrt{N_2}\psi_{\min} = \sqrt{NM}(B_1 + \sqrt{b\ln n}),$$

то есть норма среднего значения статистики $\vec{S}_N/\sqrt{NM}$ существенно больше границы критерия. Тогда с учетом (17) вероятность события

$$\left\{|\vec{S}_N| < B_1\sqrt{NM}\right\} \subset \left\{|\vec{S}_N - N\vec{\psi}| \geq \sqrt{NM}\sqrt{b\ln n}\right\}$$

есть $o(1/n)$ равномерно по $i \in \Theta$. Отсюда с учетом оценки (15) окончательно получаем

$$\begin{aligned}&\mathbf{P}\{Add(1) \neq \Theta\} \leq \\ &\leq \sum_{i\notin\Theta}\mathbf{P}\left\{|\vec{S}_N(i,\varnothing)| \geq B_1\sqrt{NM}\right\} + \sum_{i\in\Theta}\mathbf{P}\left\{|\vec{S}_N(i,\varnothing)| \geq B_1\sqrt{NM}\right\} = \\ &= o(1).\end{aligned}$$

Теорема 2 доказана. □

При выводе выражений для ранее определенного порога объема материала ограничимся сравнительно простым случаем, когда нормы всех строк $\vec{\Psi}_{\{\theta_l\}}$, $1 \leq l \leq n$, асимптотически равны. Тогда

$$\psi_{\min} \sim \psi_{\max} \sim \psi_1(\vec{f}, \vec{\varepsilon}), \tag{18}$$

где $\psi_1(\vec{f}, \vec{\varepsilon}) > 0$ — некоторая функция, определенная в каждой серии.

Равенство норм таких строк можно назвать условием *спектральной однородности совместного распределения входа и выхода* (на одноэлементных наборах существенных входов).

Для асимптотического сближения выражений для N_1, N_2 теоремы 2 потребуется также условие

$$\ln\tfrac{L-n}{\alpha}/\ln n \to \infty, \tag{19}$$

означающее согласно (14), что B_1 вносит главный вклад в соответствующие разность и сумму.

Следствие 2.1. *Пусть при условиях* **ВФ** *в схеме серий выполняются условие асимптотической спектральной однородности* (18) *и ограничения* (13), (19). *Тогда функция*

$$N_{\text{th}} = \frac{2(2^m-1)\ln(4(L-n)/\alpha)}{\psi_1^2(\vec{f}, \vec{\varepsilon})}$$

является порогом объема материала для события $\Theta^* = \Theta$.

Доказательство. Из (13) имеем $\gamma \to 0$, и тогда согласно (14)

$$B_1^2 \sim 2\ln(4(L-n)/\alpha) \sim 2\ln(4(L-n)/\alpha).$$

По условию (19), последняя величина растет быстрее $\ln n$, откуда

$$B_1/\sqrt{\ln n} \to \infty, \quad N_1 \sim N_{\text{th}} \sim N_2.$$

Кроме того, $B_1 > \sqrt{b \ln n}$ для любой константы $b > 2$ начиная с некоторого номера серии.

Пусть $\overline{\lim}\, N/N_{\text{th}} < 1$. Тогда $\overline{\lim}\, N/N_1 < 1$ и, начиная с некоторого номера серии, выполнены условия п.1 теоремы 2, откуда $\mathbf{P}\{\Theta^* = \Theta\} \to 0$.

Если же $\underline{\lim}\, N/N_{\text{th}} > 1$, то $\underline{\lim}\, N/N_2 > 1$ и, начиная с некоторой серии, выполнены условия п.2 теоремы 2, откуда $\mathbf{P}\{\Theta^* = \Theta\} \to 1$. □

Заметим, что последнее условие из (13) ограничивает область значений объема материала, в которой производится поиск порога: порядок роста N должен быть больше, чем n и $\ln((L-n)/\alpha)$. При условии (19) полученное значение N_{th} попадает в эту область, если, например, величина $(2^m - 1)/\psi_1^2(\vec{f}, \vec{\varepsilon})$ растет не медленнее чем $n/\ln n$.

Отметим также, что N_{th} логарифмически зависит от L и поэтому медленно растет с ростом L.

4. Явные выражения для порогов объема материала

В этом разделе будет явно вычислен знаменатель в выражении из следствия 2.1 для порога при наложении двух дополнительных условий:

а) равенства значений х.ф. шума при всех непустых J;

б) равенства норм всех строк спектральной матрицы Φ, соответствующих одноэлементным множествам аргументов.

Также будут приведены примеры шума и функций, для которых эти условия выполняются.

Спектр распределения двоичного случайного вектора $\vec{\xi}$ размерности M будем называть *δ-однородным*, $0 < \delta \leq 1$, если

$$|\phi_{\vec{\xi}}(J)| = \delta \text{ для всех } \varnothing \neq J \subset \{1, \ldots, M\}.$$

Например, спектр любой одномерной случайной величины ξ является δ-однородным, $\delta = \left|\mathbf{E}(-1)^{\xi}\right|$. Спектр константы $\vec{\xi} \equiv \vec{0}$ является 1-однородным.

Будем называть функцию $\vec{f}$ *спектрально 1-однородной*, если нормы n строк $\vec{\Phi}_{\{i\}}$, $1 \le i \le n$, одинаковы. Для таких функций обозначим

$$\delta^2(1,\vec{f}) = \|\vec{\Phi}_{\{i\}}\|^2 = \sum_{\varnothing \ne J \subset \{1,\dots,m\}} \phi^2_{\vec{x}\vec{f}}(\{i\},J), \quad 1 \le i \le n.$$

При δ-однородном шуме $\vec{\varepsilon}$ для спектрально 1-однородных функций согласно (12) имеем

$$\psi_{\max} = \psi_{\min} = \delta(1,\vec{f})\delta,$$

что обеспечивает выполнение условия асимптотической спектральной однородности (18).

Оставшиеся асимптотические ограничения (13), (19) следствия 2.1 для удобства объединим:

$$n \to \infty,\ \alpha \to 0,\ \left(n + \ln \tfrac{L-n}{\alpha}\right)/N \to 0,\ \ln \tfrac{L-n}{\alpha}/\ln n \to \infty. \qquad (20)$$

4.1. Векторные бент-функции

Сначала рассмотрим хорошо известный класс бент-функций. В силу отмеченной связи (2) спектра функции и ее коэффициентов Уолша-Адамара, $\vec{f}$ является бент-функцией [5, с.29] тогда и только тогда, когда модуль любого элемента ее спектральной матрицы $\Phi(\vec{f})$ равен $2^{-n/2}$. Это влечет равенство норм всех ее строк и, в частности, спектральную 1-однородность. При этом

$$\delta^2(1,\vec{f}) = (2^m - 1)2^{-n}.$$

В [6] было установлено, что такие функции существуют тогда и только тогда, когда n четное, $m \le n/2$. С учетом вышесказанного из следствия 2.1 получаем

Следствие 2.2. *Пусть при условиях* **ВФ** *в схеме серий выполняется условие* (20)*, в каждой серии n четное, $m \le n/2$ – произвольное, $\vec{f}$ – бент-функция, $\vec{\varepsilon}$ – δ-однородный шум. Тогда функция*

$$N_{\text{bent}} = \frac{2^{n+1}\ln(4(L-n)/\alpha)}{\delta^2}$$

является порогом объема материала для события $\Theta^* = \Theta$.

Интересно, что порог объема здесь не зависит от m. Это объясняется тем, что квадрат нормы каждой строки пропорционален величине $(2^m - 1)$.

Заметим также, что при равномерной мере на аргументах $\vec{z}$ бент-функции $g : \mathbb{Z}_2^{2k} \to \mathbb{Z}_2^m$ ее значения $g(\vec{z})$ имеют δ-однородный спектр, $\delta = 2^{-k}$. Такая конструкция является содержательным примером δ-однородного шума. Если $\vec{z}$ не зависит от $\vec{x}$, то и шум не будет зависеть от $\vec{x}$, что является одним из условий в **ВФ**.

4.2. Биективные двоичные отображения

Покажем, что любая биективная двоичная вектор-функция является спектрально 1-однородной. Этот факт вытекает из п.в) следующей теоремы о свойствах матрицы Φ биективного отображения $\vec{f}$.

Теорема 3. *Для $\vec{f} : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ следующие условия равносильны:*
а) $\vec{f}$ биективна;
б) $\vec{\Phi}_\varnothing(\vec{f}) = \vec{0}$.
При этом спектральная матрица обратного отображения $\vec{g}$ равна $\vec{\Phi}(\vec{g}) = \vec{\Phi}^T(\vec{f})$, и выполнено условие
в) $|\vec{\Phi}_I(\vec{f})| = 1$ для всех $\varnothing \neq I \subset \{1, \ldots, n\}$.

Доказательство. От исходной функции $\vec{f}$ перейдем к ее действительнозначному представлению $\vec{F} : \mathcal{O}_2^n \to \mathcal{O}_2^n$, где $\mathcal{O}_2 = \{\pm 1\}$ – множество корней степени 2 из единицы. Функция $\vec{F}$ однозначно задается равенством

$$\vec{F}(X) = \left((-1)^{f_1(\vec{x})}, \ldots, (-1)^{f_n(\vec{x})}\right), \; X = \left((-1)^{x_1}, \ldots, (-1)^{x_n}\right), \vec{x} \in \mathbb{Z}_2^n.$$

Считая далее, что $\vec{x} \sim U(\mathbb{Z}_2^n)$, получаем $X \sim U(\mathcal{O}_2^n)$.

1. Докажем, что пункты а) и б) эквивалентны. Легко видеть, что эквивалентны следующие условия: $\vec{f}$ биективна; $\vec{F}$ биективна; случайный вектор $Y = \vec{F}(X)$ равномерно распределен на $\mathcal{O}_2^n$.

Обозначим $X^I = X_{i_1} \ldots X_{i_r}$ для $\varnothing \neq I = \{i_1, \ldots, i_r\} \subset \{1, \ldots, n\}$, $X^\varnothing = 1$. Тогда

$$\Phi_{IJ} = \phi_{\vec{x}\vec{f}}(I, J) = \mathbf{E} X^I Y^J, \quad \Phi_{\varnothing J} = \mathbf{E} Y^J. \tag{21}$$

Согласно спектральному критерию равномерности распределения, $Y \sim U(\mathcal{O}_2^n)$ тогда и только тогда, когда $\mathbf{E} Y^J = 0$ при всех $J \neq \varnothing$, что с учетом (21) означает $\vec{\Phi}_\varnothing = \vec{0}$.

2. Осталось доказать а)⇒в). Рассмотрим пространство действительнозначных случайных величин $\xi = \xi(X)$ (действительнозначных

функций от $X \sim U(\mathcal{O}_2^n)$) со скалярным произведением $\langle \xi_1, \xi_2 \rangle = \mathbf{E}(\xi_1(X)\xi_2(X))$. Функции $X^I, I \subset \{1, \ldots, n\}$, образуют ортонормированный базис Уолша этого евклидова пространства, и равенство

$$Y^J = \vec{F}(X)^J = \sum_{I \subset \{1,\ldots,n\}} \Phi_{IJ} x^I, \quad \vec{x} \in \mathcal{O}_2^n. \tag{22}$$

является разложением в ряд функции $\vec{F}(X)^J$ по этому базису.

Пусть $X = \vec{G}(Y)$ – обратная к $\vec{F}(X)$ функция, $\vec{g}(\vec{y})$ – ее булев прообраз. В силу условия $Y \sim U(\mathcal{O}_2^n)$ система $\{y^I\}$ также ортонормирована, и аналогично (22) справедливо разложение

$$X(Y)^J = \vec{G}(Y)^J = \sum_{I \subset \{1,\ldots,n\}} c_{IJ} y^I,$$

где $c_{IJ} = \mathbf{E}X^J Y^I$. Но эти величины равны Φ_{JI} при всех $I \neq \varnothing$, а $c_{\varnothing J} = 0$ в силу спектрального критерия равномерности распределения X.

Таким образом, строки Φ суть столбцы спектральной матрицы функции $\vec{g}(\vec{y})$, норма которых равна 1 в силу равенства Парсеваля. □

Было бы интересно установить, является ли условие в) теоремы достаточным для условий а) и б).

Из п.в) теоремы 3 следует, что любая биективная двоичная вектор-функция является спектрально 1-однородной, $\delta^2(1, \vec{f}) = 1$. Тогда из следствия 2.1 получаем

Следствие 3.1. *Пусть при условиях* **ВФ** *в схеме серий выполняется условие* (20), $\vec{f} : \mathbb{Z}_2^n \to \mathbb{Z}_2^n$ *— биективная функция,* $\vec{\varepsilon}$ *— δ-однородный шум. Тогда функция*

$$N_{\text{bij}} = \frac{2(2^n - 1)\ln(4(L-n)/\alpha)}{\delta^2}$$

является порогом объема материала для события $\Theta^* = \Theta$.

Заметим, что $N_{\text{bij}} < N_{\text{bent}}$, но эти величины эквивалентны, и функция N_{bent} также является порогом объема материала при работе алгоритма на биекциях.

4.3. Булевы функции голосования

Далее рассматриваем только булевы функции $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$, то есть переходим к одномерному случаю $m = 1$. Здесь для краткости обозначаем

$$\phi_{\vec{xf}}(I) = \phi_{\vec{xf}}(I, \{1\}),\ I \subset \{1, \ldots, n\}.$$

Одномерный неравномерно распределенный шум ε всегда является δ-однородным, где

$$\delta := |\mathbf{E}(-1)^{\varepsilon}| \in (0, 1]$$

— модуль преобладания нуля в его распределении. Функция f является спектрально 1-однородной в том и только том случае, когда

$$|\phi_{\vec{xf}}\{1\}| = \ldots = |\phi_{\vec{xf}}\{n\}|.$$

Для таких функций обозначаем это значение через $\delta(1, f)$. Условие спектральной 1-однородности выполнено, в частности, для симметрических функций.

Хорошо известен класс симметрических пороговых функций

$$f(\vec{x}) = \mathbb{I}\{x_1 + \ldots + x_n \geq b\}, \quad 0 \leq b \leq n. \tag{23}$$

Их также называют *функциями голосования* (ballot functions). Величину b будем называть *параметром* функции голосования.

Если $b = b(n) = \begin{cases} n/2 - 1 \text{ или } n/2 \text{ при четном } n \\ (n-1)/2 \text{ при нечетном } n \end{cases}$, то функция голосования называется *мажоритарной* (majority). В [7, с.187] для мажоритарной функции методами теории функций комплексного переменного найдены точные значения $\phi_{\vec{xf}}(I)$, $|I| = r \geq 1$. Здесь более простыми методами найдем эти значения при $r = 1, 2$ для всех значений b.

Теорема 4. *1. Для функций голосования* (23)

$$\delta(1, f) = \phi_{\vec{xf}}\{i\} = \tfrac{1}{2^{n-1}}\tbinom{n-1}{b-1}, \quad 1 \leq i \leq n,$$
$$\phi_{\vec{xf}}\{i, j\} = \tfrac{1}{2^{n-1}}\left(\tbinom{n-2}{b-1} - \tbinom{n-2}{b-2}\right), \quad 1 \leq i < j \leq n.$$

2. Если $n \to \infty$ *и* $b(n) = \frac{n}{2} + o(n^{2/3})$, *то*

$$\delta(1, f) \sim \sqrt{\tfrac{2}{\pi n}} \exp\left(-\tfrac{2(b(n) - n/2)^2}{n}\right).$$

Доказательство. Обозначая $S(n,b)=\binom{n}{0}+\ldots+\binom{n}{b}$ и считая $\vec{x}\sim U(\mathbb{Z}_2^n)$, находим преобладание нуля в распределении значения функции

$$\mathbf{d}(f(\vec{x}))=2\mathbf{P}\{f(\vec{x})=0\}-1=\tfrac{1}{2^{n-1}}S(n,b-1)-1.$$

Обозначим через f_j^0, f_j^1 подфункции функции f, полученные фиксацией j-й переменной константами 0, 1. Замечая, что

$$f_1^0=\mathbb{I}\{x_2+\ldots+x_n\geq b\},\quad f_1^1=\mathbb{I}\{x_2+\ldots+x_n\geq b-1\},$$

с использованием формулы полной вероятности для математического ожидания и предыдущей формулы находим

$$\phi_{x\vec{f}}\{i\}=\phi_{x\vec{f}}\{1\}=\mathbf{d}(f(\vec{x})\oplus x_1)=\tfrac{1}{2}(\mathbf{d}(f_1^0)-\mathbf{d}(f_1^1))=$$
$$=\tfrac{1}{2}\left(\tfrac{1}{2^{n-2}}S(n-1,b-1)-1-\tfrac{1}{2^{n-1}}S(n-1,b-2)+1\right)=\tfrac{1}{2^{n-1}}\tbinom{n-1}{b-1}.$$

Аналогично находим

$$\phi_{x\vec{f}}\{i,j\}=\phi_{x\vec{f}}\{1,2\}=\mathbf{d}(f(\vec{x})\oplus x_1\oplus x_2)=\tfrac{1}{2}(\mathbf{d}(f_1^0\oplus x_2)-\mathbf{d}(f_1^1\oplus x_2))=$$
$$=\tfrac{1}{2}\left(\tfrac{1}{2^{n-2}}\tbinom{n-2}{b-1}-\tfrac{1}{2^{n-2}}\tbinom{n-2}{b-2}\right)=\tfrac{1}{2^{n-1}}\left(\tbinom{n-2}{b-1}-\tbinom{n-2}{b-2}\right).$$

Согласно локальной предельной теореме [3, следствие 3,с.98], для указанных значений $b(n)$ выполнено

$$\tbinom{n}{b(n)}/2^n\sim\tfrac{1}{\sqrt{2\pi n/4}}\exp(-\tilde{b}(n)^2/2),\quad \tilde{b}(n)=\tfrac{b(n)-n/2}{\sqrt{n/4}}.$$

Отсюда находим

$$\delta(1,f)=\tbinom{n-1}{b(n)-1}/2^{n-1}=\tfrac{b(n)}{n}\tbinom{n}{b(n)}/2^{n-1}\sim\tbinom{n}{b(n)}/2^n\sim$$
$$\sim\sqrt{\tfrac{2}{\pi n}}\exp\left(-\tfrac{2(b(n)-n/2)^2}{n}\right).$$

Теорема 4 доказана. □

Следствие 4.1. *1. Пусть при условиях* **ВФ** *в схеме серий выполняются условия* (20)*, и f — функция голосования с параметром $b=b(n)$. Тогда функция*

$$N_{\mathrm{bal}}(b)=\frac{2^{2n-1}}{\delta^2\binom{n-1}{b(n)-1}^2}\ln(4(L-n)/\alpha)$$

является порогом объема материала для события $\Theta^*=\Theta$.
2. Если дополнительно $b(n)=\frac{n}{2}+o(n^{2/3})$, то порог объема материала можно выписать в виде

$$N_{\mathrm{bal}}(b)=\tfrac{\pi n}{\delta^2}\ln\frac{4(L-n)}{\alpha}\exp\left(\frac{(2b-n)^2}{n}\right),$$

а при $b(n) = \frac{n}{2} + o(\sqrt{n})$ *в виде*

$$N_{\mathrm{maj}} = \tfrac{\pi n}{\delta^2} \ln \frac{4(L-n)}{\alpha}).$$

Доказательство. Из следствия 2.1 имеем выражение для порога

$$N_{\mathrm{th}} = \frac{2\ln(4(L-n)/\alpha)}{\delta^2\delta^2(1,f)},$$

и п.1 вместе с первой формулой п.2 следуют из теоремы 4.

При $b(n) = \frac{n}{2} + o(\sqrt{n})$ (т.е. для функций голосования, близких к мажоритарным), имеем $N_{\mathrm{bal}}(n,b) \sim N_{\mathrm{maj}}(n)$. Поэтому при таком поведении параметра функция N_{maj} также является порогом объема материала. □

Замечания.

1. Если i-й аргумент функции инвертируется, то у полученной функции f_1 соответствующий спектральный коэффициент меняет знак:

$$\phi_{\vec{xf}_1}\{i\} = -\phi_{\vec{xf}}\{i\}.$$

Поэтому следствие 4.1 справедливо для пороговых функций, полученных из функций голосования инверсией любых аргументов.

2. Так как $N_{\mathrm{bent}}/N_{\mathrm{maj}} = 2^{n+1}/(n\pi)$, то порог объема материала для функций голосования, близких к мажоритарным и полученных из них инверсией, в $2^{n+1}/(n\pi)$ раз меньше порога булевых бент-функций. Но первых функций значительно меньше: их число равно $2^n o(\sqrt{n})$ (число возможных инверсий, умноженное на число способов выбора отклонения $b(n)-n/2$), двоичный логарифм его эквивалентен n. При этом двоичный логарифм мощности числа бент-функций от n переменных не меньше величины $2^{n/2}$ (см., например, [5, с.24]).

3. Если $b(n) = \frac{n}{2} + o(n^{2/3})$, то $N_{\mathrm{bal}}(b) = N_{\mathrm{maj}} \exp(o(n^{1/3}))$, что по-прежнему значительно меньше N_{bent}.

4. Согласно свойствам биномиального распределения $Bin(n-1, 1/2)$, значение N_{maj} является асимптотически минимальным среди величин $N_{\mathrm{bal}}(b)$, $0 \le b \le n$. Если b значительно смещено в сторону 0 или n (функция близка к константе), то возможно $N_{\mathrm{bal}}(b)/N_{\mathrm{bent}} \to \infty$. Например, если $b(n) = o(\sqrt{n})$, то $\log_2 \binom{n-1}{b-1} \sim b \log_2 n$ и $\log_2(N_{\mathrm{bal}}(b)/N_{\mathrm{bent}}) \sim n$.

Автор благодарен А.М.Зубкову за ряд полезных замечаний, способствовавших улучшению статьи.

Список литературы

[1] Амбросимов А.С., "Свойства бент-функций q-значной логики над конечными полями", *Дискретная математика*, **6**:3 (1994), 50–60.

[2] Аренбаев Н.К., "О неравенствах для случайных векторов", *Теория вероятн. и примен.*, **22**:3 (1977), 585–589.

[3] Боровков А.А., *Теория вероятностей*, М.: Эдиториал УРСС, 1999, 472 с.

[4] Прохоров Ю.В., "О распространении неравенств С.Н.Бернштейна на многомерный случай", *Теория вероятн. и примен.*, **13**:2 (1968), 266–274.

[5] Токарева Н.Н., "Бент-функции: результаты и приложения. Обзор работ", *Прикладная дискретная математика*, 2009, № 1(3), 15–37.

[6] Nyberg K., "Perfect nonlinear S-boxes", *Advances in cryptology Eurocrypt-1991. LNCS*, **547** (1991), 378–386.

[7] Titsworth R.C., "Correlation properties of cyclic sequences. Thesis for the degree of doctor of Philosophy", 1962, 244 pp., available at thesis.library.caltech.edu.