

---

# ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

---

## PROBABILITY THEORY AND MATHEMATICAL STATISTICS

---

УДК 519.237.4

### АСИМПТОТИЧЕСКИЙ АНАЛИЗ СТАТИСТИЧЕСКИХ ОЦЕНОК ЭНТРОПИИ ШЕННОНА ДВОИЧНЫХ $s$ -ГРАММ

В. Ю. ПАЛУХА<sup>1), 2)</sup>, Ю. С. ХАРИН<sup>1), 2)</sup>

<sup>1)</sup>Научно-исследовательский институт прикладных проблем математики и информатики БГУ,  
пр. Независимости, 4, 220030, г. Минск, Беларусь

<sup>2)</sup>Белорусский государственный университет, пр. Независимости, 4, 220030, г. Минск, Беларусь

**Аннотация.** Найдены асимптотическое распределение вероятностей статистической оценки энтропии Шеннона  $s$ -грамм  $\hat{H}(s)$  и асимптотическое совместное распределение вероятностей статистических оценок энтропии Шеннона  $s$ - и  $(s+1)$ -грамм  $\hat{H}(s)$ ,  $\hat{H}(s+1)$  для равномерно распределенной случайной двоичной последовательности при ее растущей длине. Доказано, что с ростом значения  $s$  коэффициент корреляции статистических оценок энтропии Шеннона  $s$ - и  $(s+1)$ -грамм  $\hat{H}(s)$ ,  $\hat{H}(s+1)$  стремится к нулю. Теоретические результаты проиллюстрированы компьютерными экспериментами.

---

#### Образец цитирования:

Палуха ВЮ, Харин ЮС. Асимптотический анализ статистических оценок энтропии Шеннона двоичных  $s$ -грамм. *Журнал Белорусского государственного университета. Математика. Информатика*. 2025;2:62–74.  
EDN: FEJGGY

#### For citation:

Palukha UYu, Kharin YuS. Asymptotic analysis of the statistical estimators of Shannon entropy of binary  $s$ -tuples. *Journal of the Belarusian State University. Mathematics and Informatics*. 2025;2:62–74. Russian.  
EDN: FEJGGY

---

#### Авторы:

**Владимир Юрьевич Палуха** – кандидат физико-математических наук, доцент; заведующий научно-исследовательской лабораторией математических методов защиты информации, ученый секретарь<sup>1)</sup>, доцент кафедры математического моделирования и анализа данных факультета прикладной математики и информатики<sup>2)</sup>.

**Юрий Семенович Харин** – доктор физико-математических наук, академик НАН Беларуси, профессор; директор<sup>1)</sup>, профессор кафедры математического моделирования и анализа данных факультета прикладной математики и информатики<sup>2)</sup>.

#### Authors:

**Uladzimir Yu. Palukha**, PhD (physics and mathematics), docent; head of the research laboratory of mathematical methods of information security, scientific secretary<sup>a</sup> and associate professor at the department of mathematical modelling and data analysis, faculty of applied mathematics and computer science<sup>b</sup>.

[palukha@bsu.by](mailto:palukha@bsu.by)

<https://orcid.org/0009-0007-8474-1146>

**Yuriy S. Kharin**, doctor of science (physics and mathematics), academician of the National Academy of Sciences of Belarus, full professor; director<sup>a</sup> and professor at the department of mathematical modelling and data analysis, faculty of applied mathematics and computer science<sup>b</sup>.

[kharin@bsu.by](mailto:kharin@bsu.by)

<https://orcid.org/0000-0003-4226-2546>

**Ключевые слова:** энтропия Шеннона; двоичная последовательность; статистическая оценка; ковариация; коэффициент корреляции.

**Благодарность.** Авторы выражают признательность Н. С. Тегаке за помощь в проведении компьютерных экспериментов.

## ASYMPTOTIC ANALYSIS OF THE STATISTICAL ESTIMATORS OF SHANNON ENTROPY OF BINARY $s$ -TUPLES

U. Yu. PALUKHA<sup>a, b</sup>, Yu. S. KHARIN<sup>a, b</sup>

<sup>a</sup>Research Institute for Applied Problems of Mathematics and Informatics, Belarusian State University,  
4 Niezaliezhnasci Avenue, Minsk 220030, Belarus

<sup>b</sup>Belarusian State University, 4 Niezaliezhnasci Avenue, Minsk 220030, Belarus

Corresponding author: U. Yu. Palukha (palukha@bsu.by)

**Abstract.** The asymptotic probability distribution of the statistical estimate of Shannon entropy of  $s$ -tuples  $\hat{H}(s)$  and the asymptotic joint probability distribution of the statistical estimates of Shannon entropy of  $s$ - and  $(s+1)$ -tuples  $\hat{H}(s)$ ,  $\hat{H}(s+1)$  for a uniformly distributed random binary sequence with increasing length are found. It is proved that as the value of  $s$  increase, the correlation coefficient of the statistical estimates of Shannon entropy of  $s$ - and  $(s+1)$ -tuples  $\hat{H}(s)$ ,  $\hat{H}(s+1)$  tends to zero. The theoretical results are illustrated by computer experiments.

**Keywords:** Shannon entropy; binary sequence; statistical estimator; covariance; correlation coefficient.

**Acknowledgements.** The authors are grateful to M. S. Tegaka for assistance in performing computer experiments.

### Введение

Генераторы случайных и псевдослучайных последовательностей являются неотъемлемой частью систем защиты информации. Необходимым условием для данных генераторов выступает соответствие последовательностей, порождаемых ими, модели равномерно распределенной случайной последовательности<sup>1</sup>. Одним из методов оценки качества генераторов является энтропийный анализ их выходных последовательностей [1]. Наблюдаемая двоичная последовательность разбивается на непересекающиеся фрагменты длины  $s$  ( $s$ -граммы), по ним вычисляется статистическая оценка энтропии Шеннона  $\hat{H}(s)$ . В работе находятся асимптотическое (при увеличении длины двоичной последовательности) распределение вероятностей  $\hat{H}(s)$  и асимптотическое совместное распределение вероятностей  $\hat{H}(s)$ ,  $\hat{H}(s+1)$ . Доказывается, что с ростом значения  $s$  коэффициент корреляции статистических оценок энтропии Шеннона стремится к нулю. Теоретические результаты иллюстрируются компьютерными экспериментами.

### Материалы и методы исследования

Пусть имеется двоичная последовательность длины  $T = Ms(s+1)$ . Тогда ее можно разбить на  $M(s+1)$  непересекающихся  $s$ -грамм и  $M$  непересекающихся  $(s+1)$ -грамм. Частотные оценки вероятностей  $s$ -грамм  $\langle X_j \rangle$  и  $(s+1)$ -грамм  $\langle X'_l \rangle$  определяются соотношениями

$$\hat{p}_i(s) = \frac{1}{M(s+1)} \sum_{j=1}^{M(s+1)} I\{\langle X_j \rangle = i\} = \frac{v_i}{M(s+1)}, \quad i = 0, \dots, 2^s - 1, \quad (1)$$
$$\hat{p}_k(s+1) = \frac{1}{Ms} \sum_{l=1}^{Ms} I\{\langle X'_l \rangle = k\} = \frac{v_k}{Ms}, \quad k = 0, \dots, 2^{s+1} - 1.$$

<sup>1</sup>Криптология : учебник / Ю. С. Харин [и др.]. Минск : БГУ, 2013. 511 с. (Классическое университетское издание).

Статистические оценки  $s$ - и  $(s+1)$ -мерной энтропии Шеннона на основе частотных оценок вероятностей (1) имеют вид

$$\hat{H}(s) = - \sum_{i=0}^{2^s-1} \hat{p}_i(s) \ln \hat{p}_i(s) \in [0, s \ln 2],$$

$$\hat{H}(s+1) = - \sum_{k=0}^{2^{s+1}-1} \hat{p}_k(s+1) \ln \hat{p}_k(s+1) \in [0, (s+1) \ln 2].$$

Чтобы вычислить ковариацию статистических оценок энтропии Шеннона (2)  $\text{cov}_0\{\hat{H}(s), \hat{H}(s+1)\}$  при истинной гипотезе  $H_0$  о том, что наблюдаемая последовательность является равномерно распределенной случайной последовательностью, необходимо сначала найти  $2^{2s+1}$  ковариаций частотных оценок вероятностей  $\text{cov}_0\{\hat{p}_i(s), \hat{p}_k(s+1)\}$  для всех пар  $s$ - и  $(s+1)$ -грамм  $(i, k)$ ,  $i = 0, \dots, 2^s - 1$ ,  $k = 0, \dots, 2^{s+1} - 1$ . Введем обозначения

$$a_{ik} = E_0\{\hat{p}_i(s), \hat{p}_k(s+1)\}, \quad b_{ik} = E_0\{\hat{p}_i(s)\} E_0\{\hat{p}_k(s+1)\}.$$

Тогда для искоемых ковариаций частотных оценок вероятностей справедливо выражение

$$d_{ik} = \text{cov}_0\{\hat{p}_i(s), \hat{p}_k(s+1)\} = a_{ik} - b_{ik}.$$

Поскольку  $E_0\{\hat{p}_i(s)\} = 2^{-s}$ ,  $i = 0, \dots, 2^s - 1$ ,  $E_0\{\hat{p}_k(s+1)\} = 2^{-s-1}$ ,  $k = 0, \dots, 2^{s+1} - 1$ , имеем  $b_{ik} = 2^{-2s-1}$ ,  $i = 0, \dots, 2^s - 1$ ,  $k = 0, \dots, 2^{s+1} - 1$ . Для  $a_{ik}$  справедливо выражение

$$\begin{aligned} a_{ik} &= \frac{1}{M(s+1)} \frac{1}{Ms} \sum_{j=1}^{M(s+1)} \sum_{l=1}^{Ms} E_0\{I\{\langle X_j \rangle = i, \langle X'_l \rangle = k\}\} = \\ &= \frac{1}{M(s+1)} \frac{1}{Ms} \sum_{j=1}^{M(s+1)} \sum_{l=1}^{Ms} P_0\{\langle X_j \rangle = i, \langle X'_l \rangle = k\}. \end{aligned}$$

Значение величины

$$c_{ijkl} = P_0\{\langle X_j \rangle = i, \langle X'_l \rangle = k\},$$

входящей в выражение для  $a_{ik}$ , зависит от взаимного расположения  $s$ -граммы  $\langle X_j \rangle$  и  $(s+1)$ -граммы  $\langle X'_l \rangle$ . Если  $s$ -грамма  $\langle X_j \rangle$  и  $(s+1)$ -грамма  $\langle X'_l \rangle$  не пересекаются, т. е. не содержат общих элементов ряда  $\{x_1, \dots, x_T\}$ , то в силу их независимости при истинной гипотезе  $H_0$  имеем

$$c_{ijkl} = P_0\{\langle X_j \rangle = i\} P_0\{\langle X'_l \rangle = k\} = 2^{-s} \cdot 2^{-s-1} = 2^{-2s-1}, \quad i = 0, \dots, 2^s - 1, \quad k = 0, \dots, 2^{s+1} - 1. \quad (4)$$

Зафиксируем  $(s+1)$ -грамму  $\langle X'_l \rangle$ . Она может пересекаться только с двумя  $s$ -граммами. Пусть  $j'_l$  и  $j'_l + 1$  — их номера. Тогда  $a_{ik}$  представимо в виде

$$a_{ik} = \frac{1}{M(s+1)} \frac{1}{Ms} \sum_{l=1}^{Ms} \sum_{j \neq j'_l, j'_l+1} c_{ijkl} + \frac{1}{M(s+1)} \frac{1}{Ms} \sum_{l=1}^{Ms} (c_{ij'_l k l} + c_{i(j'_l+1) k l}). \quad (5)$$

Из выражения (4) для первого слагаемого в выражении (5) следует, что

$$\frac{1}{M(s+1)} \frac{1}{Ms} \sum_{l=1}^{Ms} \sum_{j \neq j'_l, j'_l+1} c_{ijkl} = \frac{1}{M(s+1)} \frac{1}{Ms} \sum_{l=1}^{Ms} \sum_{j \neq j'_l, j'_l+1} \frac{1}{2^{2s+1}} = \frac{M(s+1) - 2}{2^{2s+1} M(s+1)}. \quad (6)$$

Перед рассмотрением пересечений  $s$ - и  $(s+1)$ -грамм введем обозначение

$$l' = \begin{cases} l \bmod s, & \text{если } l \bmod s > 0, \\ s, & \text{если } l \bmod s = 0, \end{cases} \quad l' \in \{1, 2, \dots, s\}. \quad (7)$$

Пусть длина пересечения  $(s+1)$ -граммы  $\langle X'_l \rangle$  и  $s$ -граммы  $\langle X_{j'_l+1} \rangle$  равна  $m_l$ , а длина пересечения  $(s+1)$ -граммы  $\langle X'_l \rangle$  и  $s$ -граммы  $\langle X_{j'_l} \rangle$  равна  $s+1-m_l$ .

**Лемма 1.** Справедливо соотношение  $m_l = l'$ .

Доказательство. Взаимное расположение  $s$ - и  $(s+1)$ -грамм повторяется через каждые  $s$   $(s+1)$ -грамм. В связи с этим рассмотрим сначала первые  $s$   $(s+1)$ -грамм  $\langle X'_l \rangle$  ( $l = 1, \dots, s$ ).

С первой  $(s+1)$ -граммой  $\langle X'_1 \rangle$  пересекаются  $s$ -граммы  $\langle X_1 \rangle$  и  $\langle X_2 \rangle$ . Отметим, что  $s$ -грамма  $\langle X_1 \rangle$  покрывается  $(s+1)$ -граммой  $\langle X'_1 \rangle$  полностью, т. е.  $s+1 - m_1 = s$ , а длина пересечения  $s$ -граммы  $\langle X_2 \rangle$  и  $(s+1)$ -граммы  $\langle X'_1 \rangle$  равна  $m_1 = 1$ , т. е. при  $l = 1$  имеем  $m_l = l$ . Поскольку для  $l = 2, \dots, s$   $s$ -граммы  $\langle X_{j_l} \rangle$  и  $\langle X_{j_{l-1}+1} \rangle$  совпадают, длины пересечений  $(s+1)$ -граммы  $\langle X'_{l-1} \rangle$  с  $s$ -граммой  $\langle X_{j_{l-1}+1} \rangle$  и  $(s+1)$ -граммы  $\langle X'_l \rangle$  с  $s$ -граммой  $\langle X_{j_l} \rangle = \langle X_{j_{l-1}+1} \rangle$  в сумме дают  $s$ , т. е.  $m_{l-1} + s+1 - m_l = s$ , откуда  $m_l = m_{l-1} + 1$ . Таким образом, имеем рекуррентное соотношение с начальным значением  $m_1 = 1$ . Следовательно, для  $l = 1, \dots, s$  справедливо  $m_l = l$ , что при указанных значениях  $l$  равносильно  $m_l = l'$ .

В отношении  $(s+1)$ -грамм  $\langle X'_l \rangle$ ,  $l = s+1, 2s+1, \dots, ts+1, \dots$ , справедливы те же рассуждения, что и в отношении  $(s+1)$ -грамм  $\langle X'_l \rangle$ . Отличие состоит лишь в том, что  $m_l = 1 = l'$ . Дальнейшие рассуждения о сумме длин пересечений  $(s+1)$ -граммы  $\langle X'_{l-1} \rangle$  с  $s$ -граммой  $\langle X_{j_{l-1}+1} \rangle$  и  $(s+1)$ -граммы  $\langle X'_l \rangle$  с  $s$ -граммой  $\langle X_{j_l} \rangle = \langle X_{j_{l-1}+1} \rangle$  также справедливы, т. е. для  $l = ts+2, \dots, (t+1)s$ ,  $t \in \mathbb{N}$ ,  $m_l = m_{l-1} + 1$ . Следовательно,  $m_l = l'$  для любого  $l$ . Лемма доказана.

На рис. 1 продемонстрированы возможные типы пересечений при  $s = 3$ .

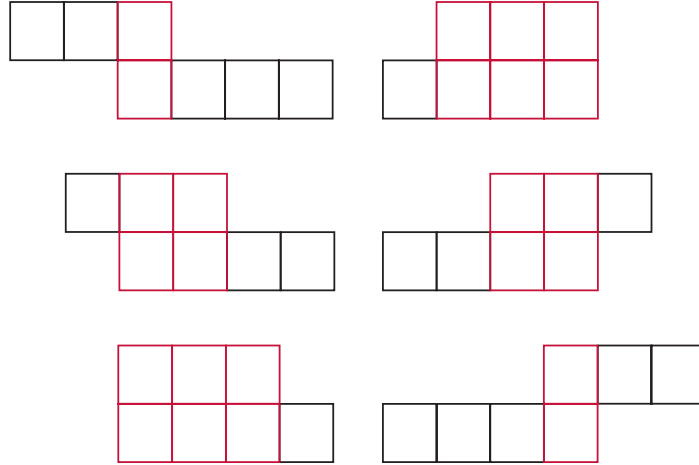


Рис. 1. Возможные типы пересечений 3- и 4-грамм

Fig. 1. Possible types of 3- and 4-tuples intersections

Полученный при пересечении  $(s+1)$ -граммы  $\langle X'_l \rangle$  и  $s$ -граммы  $\langle X_{j_l} \rangle$  вектор имеет длину  $s+1 + s - (s+1 - m_l) = s + l'$ . Соответствующий пересечению индекс  $i$  обозначим мультииндексом  $I_i^s$ , а индекс  $k$  – мультииндексом  $K_1^{s+1-l'}$ . Полученный при пересечении  $(s+1)$ -граммы  $\langle X'_l \rangle$  и  $s$ -граммы  $\langle X_{j_l+1} \rangle$  вектор имеет длину  $s+1 + s - m_l = 2s+1 - l'$ . Соответствующий пересечению индекс  $i$  обозначим мультииндексом  $I_i''$ , а индекс  $k$  – мультииндексом  $K_{s+2-l'}^{s+1}$ .

**Теорема 1.** При истинной гипотезе  $H_0$  для ковариаций (3) справедливо следующее выражение:

$$d_{ik} = \frac{1}{M(s+1)} \frac{1}{2^s} \left( \frac{1}{s} \sum_{l=1}^s \left( \frac{1}{2^l} I \{ I_i^s = K_1^{s+1-l} \} + \frac{1}{2^{s+1-l}} I \{ I_i'' = K_{s+2-l}^{s+1} \} \right) - \frac{1}{2^s} \right). \quad (8)$$

Доказательство. При расчете  $c_{ij'kl}$  и  $c_{i(j'+1)kl}$  необходимо учесть как длину двоичных векторов, полученных при пересечении, так и совпадение соответствующих пересечению элементов  $i$  и  $k$ . Тогда

$$c_{ij'kl} = P_0 \{ \langle X_{j'} \rangle = i, \langle X'_l \rangle = k \} = \frac{1}{2^{s+l'}} I \{ I_i^s = K_1^{s+1-l'} \}, \quad (9)$$

$$c_{i(j'+1)kl} = P_0 \{ \langle X_{j'+1} \rangle = i, \langle X'_l \rangle = k \} = \frac{1}{2^{2s+1-l'}} I \{ I_i'' = K_{s+2-l'}^{s+1} \}.$$

Заметим, что индикаторы в выражениях (9) представимы в виде

$$I\{I_1^s = K_1^{s+1-l'}\} = \prod_{j=1}^{s+1-l'} (i_{j'+j-1} \oplus k_j \oplus 1), \quad I\{I_1^{l'} = K_{s+2-l'}^{s+1}\} = \prod_{j=1}^{l'} (i_j \oplus k_{j+s+1-l'} \oplus 1).$$

Подставим выражения (6) и (9) в выражение (5) и получим

$$a_{ik} = \frac{M(s+1)-2}{2^{2s+1}M(s+1)} + \frac{1}{M(s+1)} \frac{1}{Ms} \sum_{l=1}^{Ms} \left( \frac{1}{2^{s+l}} I\{I_1^s = K_1^{s+1-l'}\} + \frac{1}{2^{2s+1-l'}} I\{I_1^{l'} = K_{s+2-l'}^{s+1}\} \right),$$

где зависимость  $l'$  от  $l$  определяется формулой (7). Тогда

$$\begin{aligned} d_{ik} &= \frac{M(s+1)-2}{2^{2s+1}M(s+1)} + \frac{1}{M(s+1)} \frac{1}{Ms} \sum_{l=1}^{Ms} \left( \frac{1}{2^{s+l}} I\{I_1^s = K_1^{s+1-l'}\} + \frac{1}{2^{2s+1-l'}} I\{I_1^{l'} = K_{s+2-l'}^{s+1}\} \right) - \\ &- \frac{1}{2^{2s+1}} = \frac{1}{M(s+1)} \frac{1}{Ms} \sum_{l=1}^{Ms} \left( \frac{1}{2^{s+l}} I\{I_1^s = K_1^{s+1-l'}\} + \frac{1}{2^{2s+1-l'}} I\{I_1^{l'} = K_{s+2-l'}^{s+1}\} \right) - \frac{1}{2^{2s}M(s+1)} = \\ &= \frac{1}{M(s+1)} \frac{1}{2^s} \left( \frac{1}{Ms} \sum_{l=1}^{Ms} \left( \frac{1}{2^{l'}} I\{I_1^s = K_1^{s+1-l'}\} + \frac{1}{2^{s+1-l'}} I\{I_1^{l'} = K_{s+2-l'}^{s+1}\} \right) - \frac{1}{2^s} \right). \end{aligned} \quad (10)$$

Поскольку через каждые  $s(s+1)$ -грамм слагаемые в сумме повторяются, из выражения (10) вытекает выражение (8). Теорема доказана.

**Следствие 1.** Справедлива следующая точная двусторонняя оценка ковариаций (3):

$$-\frac{1}{2^{2s}M(s+1)} \leq \text{cov}_0\{\hat{p}_i(s), \hat{p}_k(s+1)\} \leq \frac{1}{2^{s-1}s(s+1)M} \left( 1 - \frac{s+2}{2^{s+1}} \right). \quad (11)$$

**Доказательство.** В случае если индекс  $i$  состоит из одних нулей, а индекс  $k$  – из одних единиц, и наоборот, все индикаторы в выражениях (9) будут равны нулю. Тогда

$$a_{ik} = \frac{M(s+1)-2}{2^{2s+1}M(s+1)}, \quad d_{ik} = a_{ik} - b_{ik} = \frac{1}{2^{2s+1}} - \frac{1}{2^{2s}M(s+1)} - \frac{1}{2^{2s+1}} = -\frac{1}{2^{2s}M(s+1)}.$$

В случае если индексы  $i$  и  $k$  состоят из одних нулей или единиц, все индикаторы в выражениях (9) будут равны единице. Тогда

$$a_{ik} = \frac{M(s+1)-2}{2^{2s+1}M(s+1)} + \frac{1}{M(s+1)} \frac{1}{s} \sum_{l=1}^s \left( \frac{1}{2^{s+l}} + \frac{1}{2^{2s+1-l}} \right).$$

Поскольку  $s+1-l$ , как и  $l$ , пробегает все возможные значения от 1 до  $s$ , по формуле суммы геометрической прогрессии имеем

$$\sum_{l=1}^s \left( \frac{1}{2^{s+l}} + \frac{1}{2^{2s+1-l}} \right) = \frac{2}{2^s} \frac{\frac{1}{2} \left( 1 - \frac{1}{2^s} \right)}{1 - \frac{1}{2}} = \frac{1}{2^{s-1}} \left( 1 - \frac{1}{2^s} \right).$$

Окончательно получаем

$$\begin{aligned} a_{ik} &= \frac{M(s+1)-2}{2^{2s+1}M(s+1)} + \frac{1}{M(s+1)} \frac{1}{s} \frac{1}{2^{s-1}} \left( 1 - \frac{1}{2^s} \right) = \frac{M(s+1)-2}{2^{2s+1}M(s+1)} + \frac{1}{2^{s-1}sM(s+1)} \left( 1 - \frac{1}{2^s} \right), \\ d_{ik} &= a_{ik} - b_{ik} = \frac{1}{2^{2s+1}} - \frac{1}{2^{2s}M(s+1)} + \frac{1}{2^{s-1}sM(s+1)} \left( 1 - \frac{1}{2^s} \right) - \frac{1}{2^{2s+1}} = \\ &= \frac{1}{2^{s-1}sM(s+1)} \left( 1 - \frac{1}{2^s} \right) - \frac{1}{2^{2s}M(s+1)} = \frac{1}{2^{s-1}s(s+1)M} - \frac{1}{2^{2s}M(s+1)} \left( 1 + \frac{2}{s} \right) = \\ &= \frac{1}{2^{s-1}s(s+1)M} \left( 1 - \frac{s+2}{2^{s+1}} \right). \end{aligned}$$

Следствие доказано.

**Следствие 2.** Коэффициент корреляции частотных оценок вероятностей

$$\text{corr}_0\{\hat{p}_i(s), \hat{p}_k(s+1)\} = \frac{\text{cov}_0\{\hat{p}_i(s), \hat{p}_k(s+1)\}}{\sqrt{D_0\{\hat{p}_i(s)\} D_0\{\hat{p}_k(s+1)\}}} \quad (12)$$

удовлетворяет двустороннему неравенству

$$-2 \sqrt{\frac{s}{(s+1)(2^s-1)(2^{s+1}-1)}} \leq \text{corr}_0\{\hat{p}_i(s), \hat{p}_k(s+1)\} \leq \frac{2^{s+2} - 2s - 4}{\sqrt{s(s+1)(2^s-1)(2^{s+1}-1)}}. \quad (13)$$

Доказательство. Для дисперсии частотных оценок вероятностей справедливо

$$\begin{aligned} D_0\{\hat{p}_i(s)\} &= D_0\left\{\frac{1}{M(s+1)} \sum_{j=1}^{M(s+1)} I\{\langle X_j \rangle = i\}\right\} = \frac{1}{M^2(s+1)^2} \sum_{j=1}^{M(s+1)} D_0\{I\{\langle X_j \rangle = i\}\} = \\ &= \frac{1}{M^2(s+1)^2} \sum_{j=1}^{M(s+1)} \frac{1}{2^s} \left(1 - \frac{1}{2^s}\right) = \frac{1}{2^s M(s+1)} \left(1 - \frac{1}{2^s}\right), \quad i = 0, \dots, 2^s - 1. \end{aligned} \quad (14)$$

Аналогично

$$D_0\{\hat{p}_k(s+1)\} = \frac{1}{2^{s+1} M s} \left(1 - \frac{1}{2^{s+1}}\right), \quad k = 0, \dots, 2^{s+1} - 1.$$

Тогда

$$\begin{aligned} D_0\{\hat{p}_i(s)\} D_0\{\hat{p}_k(s+1)\} &= \frac{1}{2^s M(s+1)} \left(1 - \frac{1}{2^s}\right) \frac{1}{2^{s+1} M s} \left(1 - \frac{1}{2^{s+1}}\right) = \frac{(2^s - 1)(2^{s+1} - 1)}{2^{2s} \cdot 2^{2s+2} M^2 s(s+1)}, \\ \sqrt{D_0\{\hat{p}_i(s)\} D_0\{\hat{p}_k(s+1)\}} &= \frac{\sqrt{(2^s - 1)(2^{s+1} - 1)}}{2^{2s+1} M \sqrt{s(s+1)}}. \end{aligned}$$

Подставим данное выражение в формулу (12) и получим

$$\text{corr}_0\{\hat{p}_i(s), \hat{p}_k(s+1)\} = \frac{2^{2s+1} M \sqrt{s(s+1)} \text{cov}_0\{\hat{p}_i(s), \hat{p}_k(s+1)\}}{\sqrt{(2^s - 1)(2^{s+1} - 1)}}.$$

Тогда на основе двусторонней оценки (11) получим двустороннюю оценку (13). Следствие доказано.

Введем обозначения

$$f_{ik} = \sum_{l=1}^s \left( \frac{1}{2^l} I\{I_l^s = K_1^{s+1-l}\} + \frac{1}{2^{s+1-l}} I\{I_l^l = K_{s+2-l}^{s+1}\} \right), \quad g_{ik} = \frac{f_{ik}}{s} - \frac{1}{2^s}. \quad (15)$$

Тогда из выражения (8) следует, что

$$d_{ik} = \frac{g_{ik}}{2^s M(s+1)}. \quad (16)$$

**Лемма 2.** Справедливо выражение

$$\sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} f_{ik} = 2^{s+1} s.$$

Доказательство. Изменим порядок суммирования:

$$\begin{aligned} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} f_{ik} &= \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} \sum_{l=1}^s \left( \frac{1}{2^l} I\{I_l^s = K_1^{s+1-l}\} + \frac{1}{2^{s+1-l}} I\{I_l^l = K_{s+2-l}^{s+1}\} \right) = \\ &= \sum_{l=1}^s \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} \left( \frac{1}{2^l} I\{I_l^s = K_1^{s+1-l}\} + \frac{1}{2^{s+1-l}} I\{I_l^l = K_{s+2-l}^{s+1}\} \right). \end{aligned} \quad (17)$$

Стоящие внутри сумм индикаторные функции, как следует из выражений (9), соответствуют пересечениям  $s$ - и  $(s+1)$ -грамм, при этом пересечению длины  $l$  соответствует множитель  $2^{l-s-1}$  (поскольку из суммы длин  $s$ - и  $(s+1)$ -грамм была вычтена длина пересечения, а также был вынесен за скобки множитель  $2^{-s}$  в определении  $d_{ik}$ ). При суммировании по всем возможным  $s$ - и  $(s+1)$ -граммам индикаторная функция будет равна единице ровно  $2^{2s+1-l}$  раз, что соответствует числу возможных вариантов значений бит, оказавшихся вне пересечения длины  $l$ . Если рассуждать аналогично (либо переобозначить  $l$ ), то  $2^{-l}$  будет умножаться на индикаторную функцию, которая соответствует пересечению длины  $s+1-l$ . При суммировании по всем возможным  $s$ - и  $(s+1)$ -граммам индикаторная функция будет равна единице ровно  $2^{s+l}$  раз. Тогда в выражении (17) получим

$$\begin{aligned} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} f_{ik} &= \sum_{l=1}^s \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} \left( \frac{1}{2^l} I\{I_l^s = K_1^{s+1-l}\} + \frac{1}{2^{s+1-l}} I\{I_l^l = K_{s+2-l}^{s+1}\} \right) = \\ &= \sum_{l=1}^s \left( \frac{2^{s+l}}{2^l} + \frac{2^{2s+1-l}}{2^{s+1-l}} \right) = \sum_{l=1}^s (2^s + 2^s) = 2^{s+1}s. \end{aligned}$$

Лемма доказана.

**Лемма 3.** *Справедливо выражение*

$$\sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} f_{ik}^2 = 2^{s+1} + 2s^2 - s - 2. \quad (18)$$

**Доказательство.** Перепишем  $f_{ik}$  в следующем виде:

$$f_{ik} = \sum_{l=1}^s \frac{1}{2^l} (x_{ik}^{(l)} + y_{ik}^{(l)}), \quad x_{ik}^{(l)} = I\{I_l^s = K_1^{s+1-l}\}, \quad y_{ik}^{(l)} = I\{I_l^{s+1-l} = K_{l+1}^{s+1}\}.$$

Тогда для  $f_{ik}^2$  справедливо выражение

$$f_{ik}^2 = \left( \sum_{l=1}^s \frac{1}{2^l} (x_{ik}^{(l)} + y_{ik}^{(l)}) \right)^2 = \sum_{l=1}^s \frac{1}{2^{2l}} (x_{ik}^{(l)} + y_{ik}^{(l)})^2 + 2 \sum_{l=1}^{s-1} \sum_{t=l+1}^s \frac{1}{2^{l+t}} (x_{ik}^{(l)} + y_{ik}^{(l)}) (x_{ik}^{(t)} + y_{ik}^{(t)}).$$

Изменим порядок суммирования:

$$\begin{aligned} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} f_{ik}^2 &= \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} \left( \sum_{l=1}^s \frac{1}{2^{2l}} (x_{ik}^{(l)} + y_{ik}^{(l)})^2 + 2 \sum_{l=1}^{s-1} \sum_{t=l+1}^s \frac{1}{2^{l+t}} (x_{ik}^{(l)} + y_{ik}^{(l)}) (x_{ik}^{(t)} + y_{ik}^{(t)}) \right) = \\ &= \sum_{l=1}^s \frac{1}{2^{2l}} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} (x_{ik}^{(l)} + y_{ik}^{(l)})^2 + 2 \sum_{l=1}^{s-1} \sum_{t=l+1}^s \frac{1}{2^{l+t}} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} (x_{ik}^{(l)} + y_{ik}^{(l)}) (x_{ik}^{(t)} + y_{ik}^{(t)}). \end{aligned} \quad (19)$$

Справедливы следующие соотношения:

$$\begin{aligned} \frac{1}{2^{2l}} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} (x_{ik}^{(l)} + y_{ik}^{(l)})^2 &= 2^l + 1, \quad l=1, \dots, s, \\ \frac{1}{2^{l+t}} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} (x_{ik}^{(l)} + y_{ik}^{(l)}) (x_{ik}^{(t)} + y_{ik}^{(t)}) &= 2, \quad l=1, \dots, s-1, \quad t=l+1, \dots, s. \end{aligned} \quad (20)$$

Подставим выражения (20) в выражение (19) и получим

$$\begin{aligned} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} f_{ik}^2 &= \sum_{l=1}^s (2^l + 1) + 2 \sum_{l=1}^{s-1} \sum_{t=l+1}^s 2 = \sum_{l=1}^s 2^l + s + 2(s^2 - s) = \frac{2(2^s - 1)}{2 - 1} + 2s^2 - s = \\ &= 2^{s+1} - 2 + 2s^2 - s. \end{aligned}$$

Лемма доказана.

**Следствие 3.** *Справедливо выражение*

$$\sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} g_{ik}^2 = \frac{2^{s+1} - s - 2}{s^2}. \quad (21)$$

Доказательство. Из выражений (15), (18) и (21) следует, что

$$\begin{aligned} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} g_{ik}^2 &= \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} \left( \frac{f_{ik}}{s} - \frac{1}{2^s} \right)^2 = \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} \left( \frac{f_{ik}^2}{s^2} - \frac{2f_{ik}}{2^s s} + \frac{1}{2^{2s}} \right) = \frac{1}{s^2} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} f_{ik}^2 - \\ &- \frac{1}{2^{s-1}s} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} f_{ik} + \frac{2^{2s+1}}{2^{2s}} = \frac{2^{s+1} + 2s^2 - s - 2}{s^2} - \frac{2^{s+1}s}{2^{s-1}s} + 2 = \frac{2^{s+1} - s - 2}{s^2}. \end{aligned}$$

Следствие доказано.

Вернемся к статистическим оценкам (2).

**Теорема 2.** Пусть  $0 < \hat{p}_i \leq 2^{1-s}$ ,  $i = 0, \dots, 2^s - 1$ ,  $0 < \hat{p}_k \leq 2^{-s}$ ,  $k = 0, \dots, 2^{s+1} - 1$ . Тогда при истинной гипотезе  $H_0$  для ковариации и коэффициента корреляции статистических оценок энтропии Шеннона  $s$ - и  $(s+1)$ -грамм  $\hat{H}(s)$ ,  $\hat{H}(s+1)$  справедливы следующие выражения:

$$\begin{aligned} \text{cov}_0 \{ \hat{H}(s), \hat{H}(s+1) \} &= \frac{2^{s+1} - s - 2}{M^2 s^2 (s+1)^2} + o\left(\frac{1}{M^2}\right), \\ \text{corr}_0 \{ \hat{H}(s), \hat{H}(s+1) \} &= \frac{2(2^{s+1} - s - 2)}{s(s+1)\sqrt{(2^s - 1)(2^{s+1} - 1)}} + o(1). \end{aligned} \quad (22)$$

Доказательство. Обозначим случайную величину  $\xi_i = \hat{p}_i(s) - p_i$ ,  $i = 0, \dots, 2^s - 1$ . Тогда  $\hat{H}(s) = -\sum_{i=0}^{2^s-1} (p_i + \xi_i) \ln(p_i + \xi_i)$ . Воспользуемся представлением логарифма рядом Тейлора<sup>2</sup>:

$$\ln(p_i + \xi_i) = \ln\left(p_i \left(1 + \frac{\xi_i}{p_i}\right)\right) = \ln p_i + \frac{\xi_i}{p_i} - \frac{1}{2} \frac{\xi_i^2}{p_i^2} + O(|\xi_i|^3).$$

Для сходимости ряда Тейлора необходимо потребовать, чтобы  $\frac{\xi_i}{p_i} \in (-1; 1]$ , что при истинной гипотезе  $H_0$  равносильно условию  $\xi_i \in \left(-\frac{1}{2^s}; \frac{1}{2^s}\right]$ . Проверим выполнение условия  $\frac{\xi_i}{p_i} = \frac{\hat{p}_i - p_i}{p_i} = \frac{\hat{p}_i}{p_i} - 1 = 2^s \hat{p}_i - 1 \in (-1; 1]$ :

$$2^s \hat{p}_i - 1 > -1 \Leftrightarrow 2^s \hat{p}_i > 0 \Leftrightarrow \hat{p}_i > 0, \quad 2^s \hat{p}_i - 1 \leq 1 \Leftrightarrow 2^s \hat{p}_i \leq 2 \Leftrightarrow \hat{p}_i \leq \frac{2}{2^s}. \quad (23)$$

Условие (23) выполнено согласно формулировке теоремы. Тогда для статистической оценки энтропии Шеннона получим

$$\begin{aligned} \hat{H}(s) &= -\sum_{i=0}^{2^s-1} (p_i + \xi_i) \left( \ln p_i + \frac{1}{p_i} \xi_i - \frac{1}{2p_i^2} \xi_i^2 + O(|\xi_i|^3) \right) = -\sum_{i=0}^{2^s-1} p_i \ln p_i - \\ &- \sum_{i=0}^{2^s-1} \left( p_i \frac{1}{p_i} \xi_i + \xi_i \ln p_i \right) - \sum_{i=0}^{2^s-1} \left( \frac{1}{p_i} \xi_i^2 - p_i \frac{1}{2p_i^2} \xi_i^2 \right) + O\left(\max_i |\xi_i|^3\right) = \\ &= H(s) - \sum_{i=0}^{2^s-1} (1 + \ln p_i) \xi_i - \frac{1}{2} \sum_{i=0}^{2^s-1} \frac{1}{p_i} \xi_i^2 + O\left(\max_i |\xi_i|^3\right). \end{aligned}$$

При истинной гипотезе  $H_0$   $p_i = p = \frac{1}{2^s}$ ,  $i = 0, \dots, 2^s - 1$ . Кроме того,  $\sum_{i=0}^{2^s-1} \xi_i = \sum_{i=0}^{2^s-1} \hat{p}_i - \sum_{i=0}^{2^s-1} p_i = 1 - 1 = 0$ . Следовательно,

$$\begin{aligned} \hat{H}(s) &= H(s) - (1 + \ln p) \sum_{i=0}^{2^s-1} \xi_i - \frac{1}{2p} \sum_{i=0}^{2^s-1} \xi_i^2 + O\left(\max_i |\xi_i|^3\right) = \\ &= H(s) - 2^{s-1} \sum_{i=0}^{2^s-1} \xi_i^2 + O\left(\max_i |\xi_i|^3\right). \end{aligned}$$

<sup>2</sup>Воднев В. Т., Наумович А. Ф., Наумович Н. Ф. Основные математические формулы : справочник / под ред. Ю. С. Богданова. 2-е изд., перераб. и доп. Минск : Выш. шк., 1988. 272 с.

Обозначим случайную величину  $\eta_k = \hat{p}_k(s+1) - p_k$ ,  $k = 0, \dots, 2^{s+1} - 1$ . Для сходимости ряда Тейлора необходимо потребовать, чтобы  $\frac{\eta_k}{p_k} \in (-1; 1]$ . Данное условие, проверяемое аналогично условию (23), выполняется согласно формулировке теоремы. Тогда при истинной гипотезе  $H_0$  имеем  $\hat{H}(s+1) = H(s+1) - 2^s \sum_{k=0}^{2^{s+1}-1} \eta_k^2 + O\left(\max_k |\eta_k|^3\right)$ .

Введем обозначения  $u_1 = \sum_{i=0}^{2^s-1} \xi_i^2$ ,  $u_2 = \sum_{k=0}^{2^{s+1}-1} \eta_k^2$ . Тогда при истинной гипотезе  $H_0$   $\hat{H}(s) = a_1 u_1 + b_1 + v_1$ ,  $\hat{H}(s+1) = a_2 u_2 + b_2 + v_2$ , где  $a_1 = -2^{s-1}$ ;  $b_1 = H(s) = s \ln 2$ ;  $v_1 = O\left(\max_i |\xi_i|^3\right)$ ;  $a_2 = -2^s$ ;  $b_2 = H(s+1) = (s+1) \ln 2$ ;  $v_2 = O\left(\max_k |\eta_k|^3\right)$ .

Сформируем составной случайный вектор  $\zeta$  длины  $3 \cdot 2^s$  конкатенацией случайных векторов  $\xi$  и  $\eta$ :  $\zeta = (\xi \parallel \eta) \in \mathbb{R}^{2^s} \times \mathbb{R}^{2^{s+1}}$ . Тогда при  $M \rightarrow \infty$  вектор  $\zeta$  будет иметь асимптотически нормальное распределение. Приведем формулу для смешанного центрального момента 4-го порядка гауссовского вектора<sup>3</sup>

$$\sigma_{ijkl} = \sigma_{ij} \sigma_{kl} + \sigma_{ik} \sigma_{jl} + \sigma_{il} \sigma_{kj}. \quad (24)$$

Из формулы (14) для всех  $i = 0, \dots, 2^s - 1$  следует, что

$$\sigma_{ii} = E_0 \{ \xi_i^2 \} = E_0 \left\{ \left( \hat{p}_i(s) - p_i(s) \right)^2 \right\} = D_0 \{ \hat{p}_i(s) \} = \frac{1}{2^s M(s+1)} \left( 1 - \frac{1}{2^s} \right). \quad (25)$$

Аналогично для всех  $i, j = 0, \dots, 2^s - 1$ ,  $i \neq j$ , имеем

$$\begin{aligned} \sigma_{ij} &= E_0 \{ \xi_i \xi_j \} = E_0 \left\{ \left( \hat{p}_i(s) - p_i(s) \right) \left( \hat{p}_j(s) - p_j(s) \right) \right\} = \text{cov}_0 \{ \hat{p}_i(s), \hat{p}_j(s) \} = \\ &= \text{cov}_0 \left\{ \frac{v_i}{M(s+1)}, \frac{v_j}{M(s+1)} \right\} = \frac{\text{cov}_0 \{ v_i, v_j \}}{M^2(s+1)^2} = -\frac{M(s+1) p_i p_j}{M^2(s+1)^2} = -\frac{1}{2^s \cdot 2^s M(s+1)}. \end{aligned} \quad (26)$$

С учетом свойств дисперсии, формулы (24) и равенства всех  $\sigma_{ii}$ ,  $i = 0, \dots, 2^s - 1$ , при истинной гипотезе  $H_0$  получаем

$$\begin{aligned} D_0 \{ u_1 \} &= D_0 \left\{ \sum_{i=0}^{2^s-1} \xi_i^2 \right\} = E_0 \left\{ \left( \sum_{i=0}^{2^s-1} \xi_i^2 \right)^2 \right\} - E_0^2 \left\{ \sum_{i=0}^{2^s-1} \xi_i^2 \right\} = \\ &= E_0 \left\{ \sum_{i=0}^{2^s-1} \xi_i^2 \sum_{j=0}^{2^s-1} \xi_j^2 \right\} - \left( E_0 \left\{ \sum_{i=0}^{2^s-1} \xi_i^2 \right\} \right)^2 = \sum_{i=0}^{2^s-1} \sum_{j=0}^{2^s-1} E_0 \{ \xi_i^2 \xi_j^2 \} - \left( \sum_{i=0}^{2^s-1} E_0 \{ \xi_i^2 \} \right)^2 = \\ &= \sum_{i,j=0}^{2^s-1} \sigma_{ijij} - \left( \sum_{i=0}^{2^s-1} \sigma_{ii} \right)^2 = \sum_{i,j=0}^{2^s-1} (\sigma_{ii} \sigma_{jj} + 2\sigma_{ij}^2) - \left( \sum_{i=0}^{2^s-1} \sigma_{ii} \right)^2 = \\ &= \sum_{i,j=0}^{2^s-1} (\sigma_{ii}^2 + 2\sigma_{ij}^2) - \left( \sum_{i=0}^{2^s-1} \sigma_{ii} \right)^2 = 2 \sum_{i,j=0}^{2^s-1} \sigma_{ij}^2 + 2^{2s} \sigma_{ii}^2 - (2^s \sigma_{ii})^2 = \\ &= 2 \sum_{i,j=0}^{2^s-1} \sigma_{ij}^2 = 2 \left( \sum_{i=0}^{2^s-1} \sigma_{ii}^2 + \sum_{i=0}^{2^s-1} \sum_{j \neq i}^{2^s-1} \sigma_{ij}^2 \right). \end{aligned} \quad (27)$$

<sup>3</sup>Харин Ю. С., Зуев Н. М., Жук Е. Е. Теория вероятностей, математическая и прикладная статистика : учебник. Минск : БГУ, 2011. 464 с. (Классическое университетское издание).

Подставим формулы (25) и (26) в выражение (27) и получим

$$D_0\{u_1\} = 2 \left( 2^s \frac{2^{-2s}(1-2^{-s})^2}{M^2(s+1)^2} + 2^s(2^s-1) \frac{2^{-2s} \cdot 2^{-2s}}{M^2(s+1)^2} \right) = \frac{1}{2^{s-1}M^2(s+1)^2} \times \\ \times \left( (1-2^{-s})^2 + (2^s-1)2^{-2s} \right) = \frac{1}{2^{s-1}M^2(s+1)^2} (1-2^{-s+1}+2^{-s}) = \frac{2^s-1}{2^{2s-1}M^2(s+1)^2}. \quad (28)$$

Аналогично  $D_0\{u_2\} = \frac{2^{s+1}-1}{2^{2s+1}M^2s^2}$ .

Воспользовавшись формулой (24) для вычисления  $E_0\{u_1u_2\}$ , получим

$$E_0\{u_1u_2\} = E_0 \left\{ \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} \xi_i^2 \eta_k^2 \right\} = \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} E_0\{\xi_i^2 \eta_k^2\} = \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} \sigma_{iikk}. \quad (29)$$

При  $j=i, l=k$  формула (24) принимает вид  $\sigma_{iikk} = \sigma_{ii}\sigma_{kk} + 2\sigma_{ik}^2$ , где  $\sigma_{ii} = D_0\{\xi_i\}$ ,  $\sigma_{kk} = D_0\{\eta_k\}$  и  $\sigma_{ik} = \text{cov}_0\{\xi_i, \eta_k\}$ ,  $i=0, \dots, 2^s-1, k=0, \dots, 2^{s+1}-1$ . Поскольку

$$E\{\xi_i\} = E\{\eta_k\} = 0, i=0, \dots, 2^s-1, k=0, \dots, 2^{s+1}-1, \quad (30)$$

имеем  $D_0\{\xi_i\} = E_0\{\xi_i^2\}$ ,  $D_0\{\eta_k\} = E_0\{\eta_k^2\}$ ,  $i=0, \dots, 2^s-1, k=0, \dots, 2^{s+1}-1$ . Используя формулы (24) и (29) для вычисления  $\text{cov}\{u_1, u_2\}$ , получим

$$\text{cov}\{u_1, u_2\} = E_0\{u_1u_2\} - E_0\{u_1\}E_0\{u_2\} = \\ = \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} (\sigma_{ii}\sigma_{kk} + 2\sigma_{ik}^2) - \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} \sigma_{ii}\sigma_{kk} = 2 \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} \sigma_{ik}^2.$$

Из выражения (30) для  $i=0, \dots, 2^s-1, k=0, \dots, 2^{s+1}-1$  следует, что

$$\text{cov}_0\{\xi_i, \eta_k\} = E_0\{\xi_i\eta_k\} = E_0\{(\hat{p}_i(s) - p_i(s))(\hat{p}_k(s+1) - p_k(s+1))\} = \\ = \text{cov}_0\{\hat{p}_i(s), \hat{p}_k(s+1)\} = d_{ik}.$$

Таким образом,

$$2^{2s-1} \text{cov}_0\{u_1, u_2\} = 2^{2s} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} d_{ik}^2 \geq 0. \quad (31)$$

Подставим выражения (16) и (21) в выражение (31) и получим

$$2^{2s-1} \text{cov}_0\{u_1, u_2\} = \frac{2^{2s}}{2^{2s}M^2(s+1)^2} \sum_{i=0}^{2^s-1} \sum_{k=0}^{2^{s+1}-1} g_{ik}^2 = \frac{2^{s+1}-s-2}{M^2s^2(s+1)^2}. \quad (32)$$

Найдем порядок  $\nu_1$  относительно  $M$ . Справедливо ограничение сверху

$$\max_i |\xi_i|^3 \leq \sum_{i=0}^{2^s-1} |\xi_i|^3.$$

Исследуем свойства величины  $\Xi = \sum_{i=0}^{2^s-1} |\xi_i|^3$ . Поскольку, как было сказано выше, величины  $\xi_i$  имеют одинаковое асимптотически нормальное распределение, моменты всех порядков и нулевое математическое

ское ожидание, то сама величина  $\Xi$  будет иметь тот же порядок, что и ее математическое ожидание. Обозначим стандартное отклонение величин  $\xi_i$  через  $\sigma = \sqrt{D\{\xi_i\}} = \sqrt{\sigma_{ii}}$ . Из выражения (27) следует, что  $\sigma \sim O\left(\frac{1}{\sqrt{M}}\right)$ . Применим формулу для абсолютного центрального момента 4-го порядка нормальной случайной величины<sup>4</sup> с учетом того, что величины  $\xi_i$  распределены асимптотически нормально и  $E\{\xi_i\} = 0, i = 0, \dots, 2^s - 1$ . Получим  $E\{\xi_i^4\} = 3\sigma_{ii}^2 = 3\sigma^4$ . Тогда из неравенства Ляпунова<sup>5</sup> следует, что

$$\left(E\{|\xi_i|^3\}\right)^{\frac{1}{3}} \leq \left(E\{\xi_i^4\}\right)^{\frac{1}{4}} = \left(3\sigma^4\right)^{\frac{1}{4}} = 3^{\frac{1}{4}}\sigma,$$

откуда

$$E\{|\xi_i|^3\} \leq \left(E\{\xi_i^4\}\right)^{\frac{3}{4}} = 3^{\frac{3}{4}}\sigma^3 \sim O\left(\frac{1}{M^{\frac{3}{2}}}\right),$$

и при фиксированном значении  $s$  справедливо  $\Xi \sim O\left(\frac{1}{M^{\frac{3}{2}}}\right)$ , откуда  $v_1 \sim O\left(\frac{1}{M^{\frac{3}{2}}}\right)$ .

Из выражения (28) следует, что  $u_1 \sim O\left(\frac{1}{M}\right)$ . Аналогично  $u_2 \sim O\left(\frac{1}{M}\right)$ ,  $v_2 \sim O\left(\frac{1}{M^{\frac{3}{2}}}\right)$ . Значит, при вы-

числении ковариации остаточными членами  $v_1, v_2$  можно пренебречь. Учитывая свойство ковариации и применяя выражение (32), получаем

$$\begin{aligned} \text{cov}_0\{\hat{H}(s), \hat{H}(s+1)\} &= a_1 a_2 \text{cov}_0\{u_1, u_2\} + a_1 \text{cov}_0\{u_1, v_2\} + a_2 \text{cov}_0\{v_1, u_2\} + \\ &+ \text{cov}_0\{v_1, v_2\} = 2^{2s-1} \text{cov}_0\{u_1, u_2\} + O\left(\frac{1}{M^{\frac{5}{2}}}\right) = \frac{2^{s+1} - s - 2}{M^2 s^2 (s+1)^2} + o\left(\frac{1}{M^2}\right). \end{aligned}$$

Для того чтобы найти коэффициент корреляции статистических оценок энтропии Шеннона

$$\text{corr}_0\{\hat{H}(s), \hat{H}(s+1)\} = \frac{\text{cov}_0\{\hat{H}(s), \hat{H}(s+1)\}}{\sqrt{D_0\{\hat{H}(s)\} D_0\{\hat{H}(s+1)\}}}, \text{ необходимо вычислить } D_0\{\hat{H}(s)\} \text{ и } D_0\{\hat{H}(s+1)\}.$$

Аналогично после пренебрежения остаточными членами получаем

$$\begin{aligned} D_0\{\hat{H}(s)\} &= a_1^2 D_0\{u_1\} + D_0\{v_1\} + a_1 \text{cov}_0\{u_1, v_1\} = 2^{2(s-1)} D_0\{u_1\} + O\left(\frac{1}{M^{\frac{5}{2}}}\right) = \\ &= \frac{2^{2(s-1)}(2^s - 1)}{2^{2s-1} M^2 (s+1)^2} + o\left(\frac{1}{M^2}\right) = \frac{2^s - 1}{2 M^2 (s+1)^2} + o\left(\frac{1}{M^2}\right) = \frac{b_1}{M^2} (1 + o_1(1)), \\ D_0\{\hat{H}(s+1)\} &= a_2^2 D_0\{u_2\} + D_0\{v_2\} + a_2 \text{cov}_0\{u_2, v_2\} = 2^{2s} D_0\{u_2\} + O\left(\frac{1}{M^{\frac{5}{2}}}\right) = \\ &= \frac{2^{2s}(2^{s+1} - 1)}{2^{2s+1} M^2 s^2} + o\left(\frac{1}{M^2}\right) = \frac{2^{s+1} - 1}{2 M^2 s^2} + o\left(\frac{1}{M^2}\right) = \frac{b_2}{M^2} (1 + o_2(1)), \end{aligned}$$

<sup>4</sup>Харин Ю. С., Зуев Н. М., Жук Е. Е. Теория вероятностей... 464 с.

<sup>5</sup>Там же.

где  $b_1 = \frac{2^s - 1}{2(s+1)^2}$ ;  $b_2 = \frac{2^{s+1} - 1}{2s^2}$ . Применив разложение в ряд Тейлора<sup>6</sup>  $\frac{1}{\sqrt{1+x}} = 1 - \frac{1}{2}x + o(x)$ , получим

$$\begin{aligned} \frac{1}{\sqrt{D_0\{\hat{H}(s)\}D_0\{\hat{H}(s+1)\}}} &= \frac{1}{\sqrt{\left(\frac{b_1}{M^2}(1+o_1(1))\right)\left(\frac{b_2}{M^2}(1+o_2(1))\right)}} = \\ &= \frac{M^2}{\sqrt{b_1b_2}} \frac{1}{\sqrt{(1+o_1(1))(1+o_2(1))}} = \frac{M^2}{\sqrt{b_1b_2}} \left(1 - \frac{1}{2}\max\{o_1(1), o_2(1)\}\right) = \frac{M^2}{\sqrt{b_1b_2}}(1+o(1)) = \\ &= \frac{2M^2s(s+1)}{\sqrt{(2^s-1)(2^{s+1}-1)}} + o(M^2). \end{aligned}$$

Следовательно,

$$\begin{aligned} \text{corr}_0\{\hat{H}(s), \hat{H}(s+1)\} &= \left(\frac{2^{s+1}-s-2}{M^2(s+1)^2s^2} + o\left(\frac{1}{M^2}\right)\right) \left(\frac{2M^2s(s+1)}{\sqrt{(2^s-1)(2^{s+1}-1)}} + o(M^2)\right) = \\ &= \frac{2(2^{s+1}-s-2)}{s(s+1)\sqrt{(2^s-1)(2^{s+1}-1)}} + o(1). \end{aligned}$$

Заметим, что для главного члена коэффициента корреляции статистических оценок энтропии Шеннона справедливо

$$\frac{2(2^{s+1}-s-2)}{s(s+1)\sqrt{(2^s-1)(2^{s+1}-1)}} = \frac{2^{\frac{3}{2}}}{s(s+1)} \frac{1 - \frac{s+2}{2^{s+1}}}{\sqrt{(1-2^{-s})(1-2^{-s-1})}} = O\left(\frac{1}{s^2}\right).$$

Теорема доказана.

**Следствие 4.** Из формул (22) следует, что с ростом значения  $s$  коэффициент корреляции статистических оценок энтропии Шеннона  $s$ - и  $(s+1)$ -грамм стремится к нулю.

### Результаты и их обсуждение

Для демонстрации справедливости формул (22) и следствия 4 была проведена серия компьютерных экспериментов. Рассматривалась псевдослучайная последовательность, полученная алгоритмом BelT<sup>7</sup> в режиме счетчика. Для  $s = 2, \dots, 10$  из наблюдаемой последовательности брались  $K = 1000$  фрагментов длины  $T = Ms(s+1)$  с фиксированным значением  $M = 10\,000$ . По полученным  $K$  фрагментам для каждого значения  $s$  вычислялись статистические оценки энтропии Шеннона, по которым затем рассчитывались выборочные значения ковариации и коэффициента корреляции. Также для указанных значений параметров были определены теоретические значения ковариации и коэффициента корреляции статистических оценок энтропии Шеннона по главным членам в формулах (22). Проведено сравнение выборочных и теоретических значений ковариации (рис. 2) и коэффициента корреляции (рис. 3) статистических оценок энтропии Шеннона, вычисленных по  $s$ - и  $(s+1)$ -граммам, в зависимости от значения  $s$ . Как видно из рис. 2 и 3, экспериментально полученные значения ковариации и коэффициента корреляции близки к теоретическим значениям, коэффициент корреляции стремится к нулю с ростом значения  $s$ .

Следствие 4 обосновывает применение энтропийного профиля [1] для статистического тестирования криптографических генераторов: для принятия решения о справедливости гипотезы о равномерной распределенности наблюдаемой последовательности необходимо вычисление оценки энтропии при различных значениях длины фрагмента  $s$ .

<sup>6</sup>Воднев В. Т., Наумович А. Ф., Наумович Н. Ф. Основные математические формулы... 272 с.

<sup>7</sup>Информационные технологии и безопасность. Криптографические алгоритмы генерации псевдослучайных чисел : СТБ 34.101.47-2017. Введ. 09.01.2017. Минск : Госстандарт, 2017. III, 21 с.

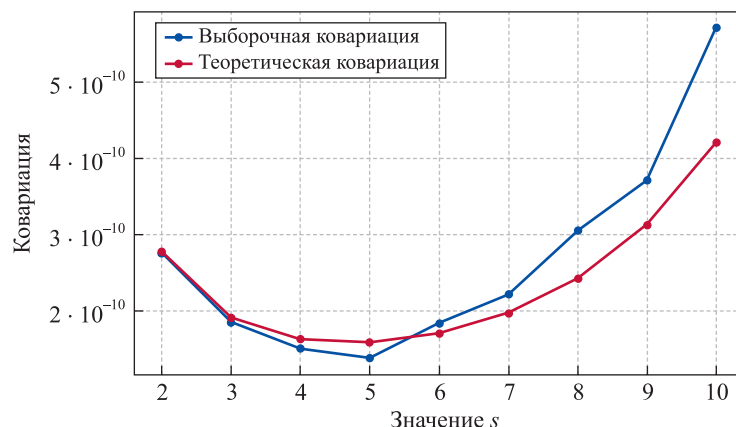


Рис. 2. Сравнение выборочных и теоретических значений ковариации статистических оценок энтропии Шеннона, вычисленных по  $s$ - и  $(s + 1)$ -граммам, в зависимости от значения  $s$

Fig. 2. Comparison of sample and theoretical values of covariance of the statistical estimates of Shannon entropy, calculated from  $s$ - and  $(s + 1)$ -tuples, depending on the value of  $s$

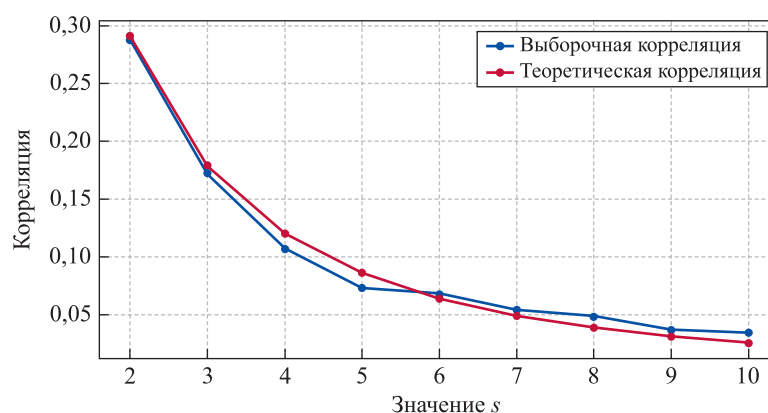


Рис. 3. Сравнение выборочных и теоретических значений коэффициента корреляции статистических оценок энтропии Шеннона, вычисленных по  $s$ - и  $(s + 1)$ -граммам, в зависимости от значения  $s$

Fig. 3. Comparison of sample and theoretical values of correlation coefficient of the statistical estimates of Shannon entropy, calculated from  $s$ - and  $(s + 1)$ -tuples, depending on the value of  $s$

## Заключение

Найдены асимптотическое (при увеличении длины двоичной последовательности) совместное распределение вероятностей частотных оценок вероятностей  $s$ - и  $(s + 1)$ -грамм, асимптотическое распределение вероятностей статистической оценки энтропии Шеннона  $s$ -грамм  $\hat{H}(s)$  и асимптотическое совместное распределение вероятностей статистических оценок энтропии Шеннона  $s$ - и  $(s + 1)$ -грамм  $\hat{H}(s)$ ,  $\hat{H}(s + 1)$ . Доказано, что с ростом значений  $s$  коэффициент корреляции статистических оценок энтропии Шеннона  $s$ - и  $(s + 1)$ -грамм  $\hat{H}(s)$ ,  $\hat{H}(s + 1)$  стремится к нулю. Теоретические результаты подтверждены компьютерными экспериментами. Обосновано применение энтропийного профиля [1] для статистического тестирования криптографических генераторов.

## Библиографические ссылки

1. Palukha UYu, Kharin YuS, Siarheeu AI, Arlou AA. On statistical testing of random and pseudorandom sequences based on entropy functionals. In: Kharin YuS, editor. *Computer data analysis and modeling: stochastic and data science. Proceedings of the 13<sup>th</sup> International conference; 2022 September 6–10; Minsk, Belarus*. Minsk: Belarusian State University; 2022. p. 148–162. EDN: QBMBVS.

Получена 06.06.2025 / исправлена 13.06.2025 / принята 20.06.2025.  
Received 06.06.2025 / revised 13.06.2025 / accepted 20.06.2025.