# 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***[1), 2)]***, Ю. С. ХАРИН***[1), 2)]

[1)]*Научно-исследовательский институт прикладных проблем математики и информатики БГУ, пр. Независимости, 4, 220030, г. Минск, Беларусь*
[2)]*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[1)], доцент кафедры математического моделирования и анализа данных факультета прикладной математики и информатики[2)].
***Юрий Семенович Харин*** – доктор физико-математических наук, академик НАН Беларуси, профессор; директор[1)], профессор кафедры математического моделирования и анализа данных факультета прикладной математики и информатики[2)].

**Authors:**
***Uladzimir Yu. Palukha***, PhD (physics and mathematics), docent; head of the research laboratory of mathematical methods of information security, scientific secretary[a] and associate professor at the department of mathematical modelling and data analysis, faculty of applied mathematics and computer science[b].
*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[a] and professor at the department of mathematical modelling and data analysis, faculty of applied mathematics and computer science[b].
*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***[a, b]***, Yu. S. KHARIN***[a, b]

[a]*Research Institute for Applied Problems of Mathematics and Informatics, Belarusian State University, 4 Niezaliezhnasci Avenue, Minsk 220030, Belarus*
[b]*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[1]. 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 *Ms* непересекающихся $(s+1)$-грамм. Частотные оценки вероятностей *s*-грамм $\langle X_j \rangle$ и $(s+1)$-грамм $\langle X'_l \rangle$ определяются соотношениями

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

[1]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

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

Чтобы вычислить ковариацию статистических оценок энтропии Шеннона (2) $\operatorname{cov}_0\{\hat{H}(s), \hat{H}(s+1)\}$ 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 $\operatorname{cov}_0\{\hat{p}_i(s),\ \hat{p}_k(s+1)\}$ для всех пар $s$- и $(s+1)$-грамм $(i,\ k)$, $i = 0, \ldots, 2^s - 1$, $k = 0, \ldots, 2^{s+1} - 1$. Введем обозначения

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

Тогда для искомых ковариаций частотных оценок вероятностей справедливо выражение

$$d_{ik} = \operatorname{cov}_0\{\hat{p}_i(s),\ \hat{p}_k(s+1)\} = a_{ik} - b_{ik}. \tag{3}$$

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

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

Значение величины

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

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 $\{x_1, \ldots, x_T\}$, то в силу их независимости при истинной гипотезе $H_0$ имеем

$$c_{ikjl} = P_0\{\langle X_j \rangle = i\} P_0\{\langle X'_l \rangle = k\} = 2^{-s} \cdot 2^{-s-1} = 2^{-2s-1},\ i = 0, \ldots, 2^s - 1,\ k = 0, \ldots, 2^{s+1} - 1. \tag{4}$$

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

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

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

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

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

$$l' = \begin{cases} l \bmod s, \text{ если } l \bmod s > 0, \\ s, \text{ если } l \bmod s = 0, \end{cases} \quad l' \in \{1, 2, \ldots, s\}. \tag{7}$$

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 $\langle X'_l \rangle$ ($l = 1, \ldots, s$).

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 $l = 2, \ldots, s$ $s$-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 $l = 1, \ldots, s$ справедливо $m_l = l$, что при указанных значениях $l$ равносильно $m_l = l'$.

В отношении $(s+1)$-грамм $\langle X'_l \rangle$, $l = s+1, 2s+1, \ldots, ts+1, \ldots$, справедливы те же рассуждения, что и в отношении $(s+1)$-грамм $\langle X'_1 \rangle$. 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 $l = ts + 2, \ldots, (t+1)s$, $t \in \mathbb{N}$, $m_l = m_{l-1} + 1$. Следовательно, $m_l = l'$ для любого $l$. Лемма доказана.

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

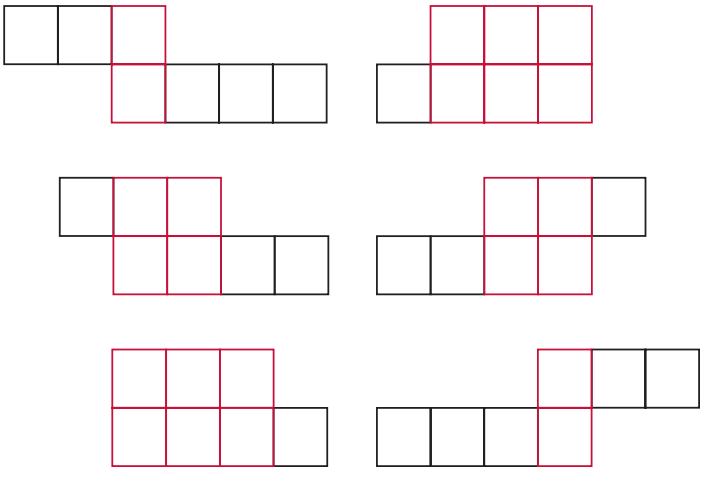

*Рис. 1.* Возможные типы пересечений 3- и 4-грамм
*Fig. 1.* Possible types of 3- and 4-tuples intersections

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 $I^s_{l'}$, 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 $I^{l'}_1$, а индекс $k$ – мультииндексом $K^{s+1}_{s+2-l'}$.

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

$$d_{ik} = \frac{1}{M(s+1)} \frac{1}{2^s} \left( \frac{1}{s} \sum_{l=1}^{s} \left( \frac{1}{2^l} I\left\{ I^s_l = K_1^{s+1-l} \right\} + \frac{1}{2^{s+1-l}} I\left\{ I^l_1 = K^{s+1}_{s+2-l} \right\} \right) - \frac{1}{2^s} \right). \tag{8}$$

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

$$\begin{aligned} c_{ij'_lkl} &= P_0\left\{ \langle X_{j'} \rangle = i, \langle X'_l \rangle = k \right\} = \frac{1}{2^{s+l'}} I\left\{ I^s_{l'} = K_1^{s+1-l'} \right\}, \\ c_{i(j'_l+1)kl} &= P_0\left\{ \langle X_{j'+1} \rangle = i, \langle X'_l \rangle = k \right\} = \frac{1}{2^{2s+1-l'}} I\left\{ I^{l'}_1 = K^{s+1}_{s+2-l'} \right\}. \end{aligned} \tag{9}$$

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

$$I\left\{I_{l'}^{s}=K_{1}^{s+1-l'}\right\}=\prod_{j=1}^{s+1-l'}\left(i_{l'+j-1}\oplus k_{j}\oplus 1\right),\ I\left\{I_{1}^{l''}=K_{s+2-l'}^{s+1}\right\}=\prod_{j=1}^{l'}\left(i_{j}\oplus k_{j+s+1-l'}\oplus 1\right).$$

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

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

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

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

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

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

$$-\frac{1}{2^{2s}M\left(s+1\right)}\le\operatorname{cov}_{0}\left\{\hat{p}_{i}\left(s\right),\ \hat{p}_{k}\left(s+1\right)\right\}\le\frac{1}{2^{s-1}s\left(s+1\right)M}\left(1-\frac{s+2}{2^{s+1}}\right).\tag{11}$$

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

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

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

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

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

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

Следствие доказано.

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

$$\operatorname{corr}_0\left\{\hat{p}_i(s),\ \hat{p}_k(s+1)\right\}=\frac{\operatorname{cov}_0\left\{\hat{p}_i(s),\ \hat{p}_k(s+1)\right\}}{\sqrt{D_0\left\{\hat{p}_i(s)\right\}D_0\left\{\hat{p}_k(s+1)\right\}}} \tag{12}$$

*удовлетворяет двустороннему неравенству*

$$-2\sqrt{\frac{s}{(s+1)\left(2^s-1\right)\left(2^{s+1}-1\right)}}\le\operatorname{corr}_0\left\{\hat{p}_i(s),\ \hat{p}_k(s+1)\right\}\le\frac{2^{s+2}-2s-4}{\sqrt{s(s+1)\left(2^s-1\right)\left(2^{s+1}-1\right)}}. \tag{13}$$

Доказательство. Для дисперсии частотных оценок вероятностей справедливо

$$D_0\left\{\hat{p}_i(s)\right\}=D_0\left\{\frac{1}{M(s+1)}\sum_{j=1}^{M(s+1)}I\left\{\left\langle X_j\right\rangle=i\right\}\right\}=\frac{1}{M^2(s+1)^2}\sum_{j=1}^{M(s+1)}D_0\left\{I\left\{\left\langle X_j\right\rangle=i\right\}\right\}=$$

$$=\frac{1}{M^2(s+1)^2}\sum_{j=1}^{M(s+1)}\frac{1}{2^s}\left(1-\frac{1}{2^s}\right)=\frac{1}{2^sM(s+1)}\left(1-\frac{1}{2^s}\right),\ i=0,\ \ldots,\ 2^s-1. \tag{14}$$

Аналогично

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

Тогда

$$D_0\left\{\hat{p}_i(s)\right\}D_0\left\{\hat{p}_k(s+1)\right\}=\frac{1}{2^sM(s+1)}\left(1-\frac{1}{2^s}\right)\frac{1}{2^{s+1}Ms}\left(1-\frac{1}{2^{s+1}}\right)=\frac{\left(2^s-1\right)\left(2^{s+1}-1\right)}{2^{2s}\cdot2^{2s+2}M^2s(s+1)},$$

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

Подставим данное выражение в формулу (12) и получим

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

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

Введем обозначения

$$f_{ik}=\sum_{l=1}^{s}\left(\frac{1}{2^l}I\left\{I_l^s=K_1^{s+1-l}\right\}+\frac{1}{2^{s+1-l}}I\left\{I_1^l=K_{s+2-l}^{s+1}\right\}\right),\ g_{ik}=\frac{f_{ik}}{s}-\frac{1}{2^s}. \tag{15}$$

Тогда из выражения (8) следует, что

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

**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:

$$\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}f_{ik}=\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}\sum_{l=1}^{s}\left(\frac{1}{2^l}I\left\{I_l^s=K_1^{s+1-l}\right\}+\frac{1}{2^{s+1-l}}I\left\{I_1^l=K_{s+2-l}^{s+1}\right\}\right)=$$

$$=\sum_{l=1}^{s}\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}\left(\frac{1}{2^l}I\left\{I_l^s=K_1^{s+1-l}\right\}+\frac{1}{2^{s+1-l}}I\left\{I_1^l=K_{s+2-l}^{s+1}\right\}\right). \tag{17}$$

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

$$\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1} f_{ik} = \sum_{l=1}^{s}\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}\left(\frac{1}{2^l} I\left\{I_l^s = K_1^{s+1-l}\right\} + \frac{1}{2^{s+1-l}} I\left\{I_1^l = K_{s+2-l}^{s+1}\right\}\right) =$$

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

Лемма доказана.

**Лемма 3.** *Справедливо выражение*

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

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

$$f_{ik} = \sum_{l=1}^{s}\frac{1}{2^l}\left(x_{ik}^{(l)} + y_{ik}^{(l)}\right),\ x_{ik}^{(l)} = I\left\{I_l^s = K_1^{s+1-l}\right\},\ y_{ik}^{(l)} = I\left\{I_1^{s+1-l} = K_{l+1}^{s+1}\right\}.$$

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

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

Изменим порядок суммирования:

$$\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1} f_{ik}^2 = \sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}\left(\sum_{l=1}^{s}\frac{1}{2^{2l}}\left(x_{ik}^{(l)} + y_{ik}^{(l)}\right)^2 + 2\sum_{l=1}^{s-1}\sum_{t=l+1}^{s}\frac{1}{2^{l+t}}\left(x_{ik}^{(l)} + y_{ik}^{(l)}\right)\left(x_{ik}^{(t)} + y_{ik}^{(t)}\right)\right) =$$

$$= \sum_{l=1}^{s}\frac{1}{2^{2l}}\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}\left(x_{ik}^{(l)} + y_{ik}^{(l)}\right)^2 + 2\sum_{l=1}^{s-1}\sum_{t=l+1}^{s}\frac{1}{2^{l+t}}\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}\left(x_{ik}^{(l)} + y_{ik}^{(l)}\right)\left(x_{ik}^{(t)} + y_{ik}^{(t)}\right). \tag{19}$$

Справедливы следующие соотношения:

$$\begin{gathered}\frac{1}{2^{2l}}\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}\left(x_{ik}^{(l)} + y_{ik}^{(l)}\right)^2 = 2^l + 1,\ l = 1, \ldots, s,\\ \frac{1}{2^{l+t}}\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}\left(x_{ik}^{(l)} + y_{ik}^{(l)}\right)\left(x_{ik}^{(t)} + y_{ik}^{(t)}\right) = 2,\ l = 1, \ldots, s-1,\ t = l+1, \ldots, s.\end{gathered} \tag{20}$$

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

$$\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1} f_{ik}^2 = \sum_{l=1}^{s}\left(2^l + 1\right) + 2\sum_{l=1}^{s-1}\sum_{t=l+1}^{s} 2 = \sum_{l=1}^{s} 2^l + s + 2\left(s^2 - s\right) = \frac{2\left(2^s - 1\right)}{2-1} + 2s^2 - s =$$

$$= 2^{s+1} - 2 + 2s^2 - s.$$

Лемма доказана.

**Следствие 3.** *Справедливо выражение*

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

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

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

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

Следствие доказано.

Вернемся к статистическим оценкам (2).

**Теорема 2.** *Пусть* $0<\hat{p}_i\le 2^{1-s}$, $i=0,\ldots,2^s-1$, $0<\hat{p}_k\le 2^{-s}$, $k=0,\ldots,2^{s+1}-1$. *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:*

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

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

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

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 $\frac{\xi_i}{p_i}=\frac{\hat{p}_i-p_i}{p_i}=\frac{\hat{p}_i}{p_i}-1=$ $=2^s\hat{p}_i-1\in(-1;1]$:

$$2^s\hat{p}_i-1>-1\Leftrightarrow 2^s\hat{p}_i>0\Leftrightarrow\hat{p}_i>0,\ 2^s\hat{p}_i-1\le 1\Leftrightarrow 2^s\hat{p}_i\le 2\Leftrightarrow\hat{p}_i\le\frac{2}{2^s}. \tag{23}$$

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

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

$$-\sum_{i=0}^{2^s-1}\left(p_i\frac{1}{p_i}\xi_i+\xi_i\ln p_i\right)-\sum_{i=0}^{2^s-1}\left(\frac{1}{p_i}\xi_i^2-p_i\frac{1}{2p_i^2}\xi_i^2\right)+O\left(\max_i|\xi_i|^3\right)=$$

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

При истинной гипотезе $H_0$ $p_i=p=\frac{1}{2^s}$, $i=0,\ldots,2^s-1$. 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,

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

$$=H(s)-2^{s-1}\sum_{i=0}^{2^s-1}\xi_i^2+O\left(\max_i|\xi_i|^3\right).$$

[2] *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 $\eta_k = \hat{p}_k(s+1) - p_k$, $k = 0, \ldots, 2^{s+1} - 1$. 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 $\hat{H}(s+1) = H(s+1) - 2^s \sum\limits_{k=0}^{2^{s+1}-1} \eta_k^2 + O\left(\max\limits_k |\eta_k|^3\right)$.

Введем обозначения $u_1 = \sum\limits_{i=0}^{2^s-1} \xi_i^2$, $u_2 = \sum\limits_{k=0}^{2^{s+1}-1} \eta_k^2$. 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 $a_1 = -2^{s-1}$; $b_1 = H(s) = s \ln 2$; $v_1 = O\left(\max\limits_i |\xi_i|^3\right)$; $a_2 = -2^s$; $b_2 = H(s+1) = (s+1)\ln 2$; $v_2 = O\left(\max\limits_k |\eta_k|^3\right)$.

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

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

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

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

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

$$\sigma_{ij} = E_0\{\xi_i \xi_j\} = E_0\left\{\left(\hat{p}_i(s) - p_i(s)\right)\left(\hat{p}_j(s) - p_j(s)\right)\right\} = \operatorname{cov}_0\{\hat{p}_i(s),\ \hat{p}_j(s)\} =$$

$$= \operatorname{cov}_0\left\{\frac{\nu_i}{M(s+1)}, \frac{\nu_j}{M(s+1)}\right\} = \frac{\operatorname{cov}_0\{\nu_i, \nu_j\}}{M^2(s+1)^2} = -\frac{M(s+1)p_i p_j}{M^2(s+1)^2} = -\frac{1}{2^s \cdot 2^s M(s+1)}. \tag{26}$$

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

$$D_0\{u_1\} = D_0\left\{\sum_{i=0}^{2^s-1} \xi_i^2\right\} = E_0\left\{\left(\sum_{i=0}^{2^s-1} \xi_i^2\right)^2\right\} - E_0^2\left\{\sum_{i=0}^{2^s-1} \xi_i^2\right\} =$$

$$= E_0\left\{\sum_{i=0}^{2^s-1} \xi_i^2 \sum_{j=0}^{2^s-1} \xi_j^2\right\} - \left(E_0\left\{\sum_{i=0}^{2^s-1} \xi_i^2\right\}\right)^2 = \sum_{i=0}^{2^s-1}\sum_{j=0}^{2^s-1} E_0\{\xi_i^2 \xi_j^2\} - \left(\sum_{i=0}^{2^s-1} E_0\{\xi_i^2\}\right)^2 =$$

$$= \sum_{i,\,j=0}^{2^s-1} \sigma_{iijj} - \left(\sum_{i=0}^{2^s-1} \sigma_{ii}\right)^2 = \sum_{i,\,j=0}^{2^s-1} \left(\sigma_{ii}\sigma_{jj} + 2\sigma_{ij}^2\right) - \left(\sum_{i=0}^{2^s-1} \sigma_{ii}\right)^2 =$$

$$= \sum_{i,\,j=0}^{2^s-1} \left(\sigma_{ii}^2 + 2\sigma_{ij}^2\right) - \left(\sum_{i=0}^{2^s-1} \sigma_{ii}\right)^2 = 2\sum_{i,\,j=0}^{2^s-1} \sigma_{ij}^2 + 2^{2s}\sigma_{ii}^2 - \left(2^s \sigma_{ii}\right)^2 =$$

$$= 2\sum_{i,\,j=0}^{2^s-1} \sigma_{ij}^2 = 2\left(\sum_{i=0}^{2^s-1} \sigma_{ii}^2 + \sum_{i=0}^{2^s-1}\sum_{j\neq i} \sigma_{ij}^2\right). \tag{27}$$

[3] *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

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

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

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

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

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

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 $\sigma_{ik} = \operatorname{cov}_0\{\xi_i, \eta_k\}$, $i = 0, \ldots, 2^s-1$, $k = 0, \ldots, 2^{s+1}-1$. Поскольку

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

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

$$\operatorname{cov}\{u_1, u_2\} = E_0\{u_1u_2\} - E_0\{u_1\}E_0\{u_2\} =$$

$$= \sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}\left(\sigma_{ii}\sigma_{kk} + 2\sigma_{ik}^2\right) - \sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}\sigma_{ii}\sigma_{kk} = 2\sum_{i=0}^{2^s-1}\sum_{k=0}^{2^{s+1}-1}\sigma_{ik}^2.$$

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

$$\operatorname{cov}_0\{\xi_i, \eta_k\} = E_0\{\xi_i\eta_k\} = E_0\left\{\left(\hat{p}_i(s) - p_i(s)\right)\left(\hat{p}_k(s+1) - p_k(s+1)\right)\right\} =$$

$$= \operatorname{cov}_0\left\{\hat{p}_i(s),\ \hat{p}_k(s+1)\right\} = d_{ik}.$$

Таким образом,

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

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

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

Найдем порядок $v_1$ 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 математиче-

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[4] с учетом того, что величины $\xi_i$ распределены асимптотически нормально и $E\{\xi_i\} = 0$, $i = 0, \ldots, 2^s - 1$. Получим $E\{\xi_i^4\} = 3\sigma_{ii}^2 = 3\sigma^4$. Тогда из неравенства Ляпунова[5] следует, что

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

откуда

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

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

$$\operatorname{cov}_0\left\{\hat{H}(s),\, \hat{H}(s+1)\right\} = a_1 a_2 \operatorname{cov}_0\{u_1, u_2\} + a_1 \operatorname{cov}_0\{u_1, v_2\} + a_2 \operatorname{cov}_0\{v_1, u_2\} +$$

$$+ \operatorname{cov}_0\{v_1, v_2\} = 2^{2s-1}\operatorname{cov}_0\{u_1, u_2\} + O\left(\frac{1}{M^{\frac{5}{2}}}\right) = \frac{2^{s+1} - s - 2}{M^2 s^2 (s+1)^2} + o\left(\frac{1}{M^2}\right).$$

Для того чтобы найти коэффициент корреляции статистических оценок энтропии Шеннона $\operatorname{corr}_0\left\{\hat{H}(s),\, \hat{H}(s+1)\right\} = \frac{\operatorname{cov}_0\left\{\hat{H}(s),\, \hat{H}(s+1)\right\}}{\sqrt{D_0\left\{\hat{H}(s)\right\} D_0\left\{\hat{H}(s+1)\right\}}}$, необходимо вычислить $D_0\left\{\hat{H}(s)\right\}$ и $D_0\left\{\hat{H}(s+1)\right\}$. Аналогично после пренебрежения остаточными членами получаем

$$D_0\left\{\hat{H}(s)\right\} = a_1^2 D_0\{u_1\} + D_0\{v_1\} + a_1 \operatorname{cov}_0\{u_1, v_1\} = 2^{2(s-1)} D_0\{u_1\} + O\left(\frac{1}{M^{\frac{5}{2}}}\right) =$$

$$= \frac{2^{2(s-1)}\left(2^s - 1\right)}{2^{2s-1} M^2 (s+1)^2} + o\left(\frac{1}{M^2}\right) = \frac{2^s - 1}{2M^2(s+1)^2} + o\left(\frac{1}{M^2}\right) = \frac{b_1}{M^2}\left(1 + o_1(1)\right),$$

$$D_0\left\{\hat{H}(s+1)\right\} = a_2^2 D_0\{u_2\} + D_0\{v_2\} + a_2 \operatorname{cov}_0\{u_2, v_2\} = 2^{2s} D_0\{u_2\} + O\left(\frac{1}{M^{\frac{5}{2}}}\right) =$$

$$= \frac{2^{2s}\left(2^{s+1} - 1\right)}{2^{2s+1} M^2 s^2} + o\left(\frac{1}{M^2}\right) = \frac{2^{s+1} - 1}{2M^2 s^2} + o\left(\frac{1}{M^2}\right) = \frac{b_2}{M^2}\left(1 + o_2(1)\right),$$

[4] *Харин Ю. С., Зуев Н. М., Жук Е. Е.* Теория вероятностей… 464 с.
[5] 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[6] $\frac{1}{\sqrt{1+x}} = 1 - \frac{1}{2}x + o(x)$, получим

$$\frac{1}{\sqrt{D_0\{\hat{H}(s)\} D_0\{\hat{H}(s+1)\}}} = \frac{1}{\sqrt{\left(\frac{b_1}{M^2}(1+o_1(1))\right)\left(\frac{b_2}{M^2}(1+o_2(1))\right)}} =$$

$$= \frac{M^2}{\sqrt{b_1 b_2}} \frac{1}{\sqrt{(1+o_1(1))(1+o_2(1))}} = \frac{M^2}{\sqrt{b_1 b_2}}\left(1 - \frac{1}{2}\max\{o_1(1), o_2(1)\}\right) = \frac{M^2}{\sqrt{b_1 b_2}}(1+o(1)) =$$

$$= \frac{2M^2 s(s+1)}{\sqrt{(2^s-1)(2^{s+1}-1)}} + o(M^2).$$

Следовательно,

$$\mathrm{corr}_0\{\hat{H}(s), \hat{H}(s+1)\} = \left(\frac{2^{s+1} - s - 2}{M^2(s+1)^2 s^2} + o\left(\frac{1}{M^2}\right)\right)\left(\frac{2M^2 s(s+1)}{\sqrt{(2^s-1)(2^{s+1}-1)}} + o(M^2)\right) =$$

$$= \frac{2(2^{s+1} - s - 2)}{s(s+1)\sqrt{(2^s-1)(2^{s+1}-1)}} + o(1).$$

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[7] в режиме счетчика. Для $s = 2, \ldots, 10$ 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 $s$.

[6] *Воднев В. Т., Наумович А. Ф., Наумович Н. Ф.* Основные математические формулы… 272 с.

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

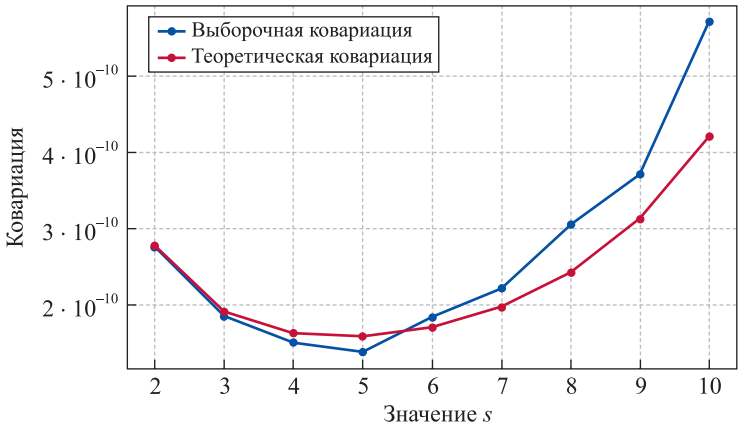


*Рис. 2.* Сравнение выборочных и теоретических значений ковариации статистических оценок энтропии Шеннона, вычисленных по $s$- и $(s+1)$-граммам, в зависимости от значения $s$

*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 $s$

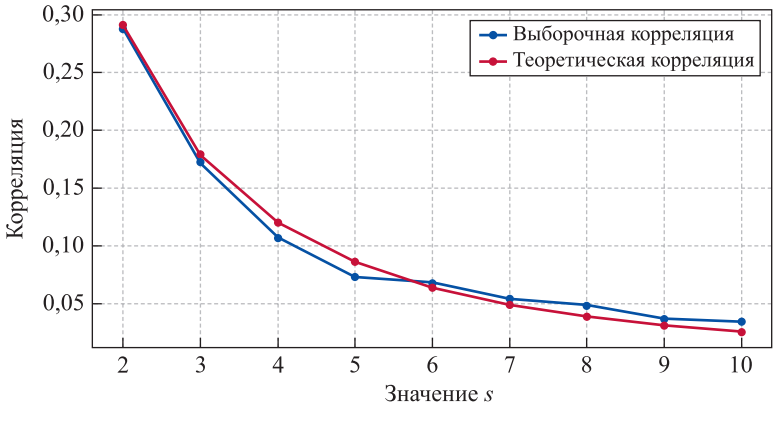


*Рис. 3.* Сравнение выборочных и теоретических значений коэффициента корреляции статистических оценок энтропии Шеннона, вычисленных по $s$- и $(s+1)$-граммам, в зависимости от значения $s$

*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: stochastics and data science. Proceedings of the 13th 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.*