

# ОБНАРУЖЕНИЕ ВКРАПЛЕНИЙ В ДВОИЧНУЮ ЦЕПЬ МАРКОВА НА ОСНОВЕ ЭНТРОПИЙНЫХ ХАРАКТЕРИСТИК

Андрей Чеславович Шимко

Научный руководитель: Егор Валентинович Вечерко

Факультет прикладной математики и информатики

Кафедра математического моделирования и анализа данных

Минск, 2017

# Математическая модель

Дана последовательность случайных величин распределенных по нормальному закону

$$\{x_t\} \text{ ЦМ}(1), x_i \in V = \{0, 1\}, i = \overline{1, T} \quad (1)$$

$$\pi = (\frac{1}{2}, \frac{1}{2}), P(\varepsilon) = \frac{1}{2}(\frac{1+\varepsilon}{1-\varepsilon}, \frac{1-\varepsilon}{1+\varepsilon}), |\varepsilon| < 1, \varepsilon \neq 0. \quad (2)$$

Сообщение

$$\mathcal{L}m_t = Bi(1, \theta), m_i \in V = \{0, 1\}, i = \overline{1, \tau}, \quad (3)$$

Ключ

$$\mathcal{L}\gamma_t = Bi(1, \delta), \gamma_i \in V = \{0, 1\}, i = \overline{1, T}; \quad (4)$$

И задано функциональное преобразование

$$y_t = \begin{cases} x_t, \gamma_t = 0; \\ m_{\tau_t}, \gamma_t = 1; \end{cases} = (1 - \gamma_t)x_t + \gamma_t m_{\tau_t}, \text{ где } \tau_t = \sum_{j=1}^t \gamma_j \quad (5)$$

Введем понятие энтропии на знак для  $l$ -граммы:

$$H_l(\delta) = -\frac{1}{l} \sum_{(a_1, \dots, a_l) \in \{0,1\}^l} P\{y_{t-l} = a_1, \dots, y_{t-1} = a_l\} \log P\{y_{t-l} = a_1, \dots, y_{t-1} = a_l\}. \quad (6)$$

## Определение

Величина

$$I\{b_i\} = -\log p_i \quad (7)$$

называется собственной информацией, содержащейся в исходе  $b_i \in B$ .

Величина  $I\{b_i\}$  изменяется от нуля в случае реализации достоверного исхода до бесконечности, когда  $p(b_i) = p_i \rightarrow 0$ . Величину  $I\{b_i\}$  можно интерпретировать как априорную неопределенность события  $\{\xi = b_i\}$ .

Случайная величина  $I\{\xi\}$  имеет математическое ожидание

$$EI\{\xi\} = -\sum_{b_i \in B} p_i \log p_i. \quad (8)$$

# Оценка энтропии биграммы

## Лемма

*Если имеет место монобитная модель вкраплений (1)-(5), то вероятности появления всевозможных биграмм имеют вид :*

$$P\{y_{t-1} = 1, y_t = 1\} = \frac{1}{4}(1 + \varepsilon)(1 - \delta)^2 + \theta\delta(1 - \delta) + \theta^2\delta^2; \quad (9)$$

$$P\{y_{t-1} = 1, y_t = 0\} = \frac{1}{4}(1 - \varepsilon)(1 - \delta)^2 + \frac{1}{2}\delta(1 - \delta) + \theta(1 - \theta)\delta^2; \quad (10)$$

$$P\{y_{t-1} = 0, y_t = 1\} = \frac{1}{4}(1 - \varepsilon)(1 - \delta)^2 + \frac{1}{2}\delta(1 - \delta) + \theta(1 - \theta)\delta^2; \quad (11)$$

$$P\{y_{t-1} = 0, y_t = 0\} = \frac{1}{4}(1 + \varepsilon)(1 - \delta)^2 + \delta(1 - \theta)(1 - \delta) + \delta^2(1 - \theta)^2. \quad (12)$$

## Лемма

*Если имеет место монобитная модель вкраплений (1)-(5), то для энтропии при  $l = 2$  справедливо асимптотическое разложение при  $\delta \rightarrow 0$  1-го порядка*

$$H_2(\delta) = H_2(0) + 2\delta\varepsilon \log \frac{1 + \varepsilon}{1 - \varepsilon} + O(\delta^2). \quad (13)$$

# Оценка энтропии 3-граммы

## Теорема

Если имеет место монобитная модель вкраплений (1)-(5), то для энтропии при  $l = 3$  справедливо асимптотическое разложение при  $\delta \rightarrow 0$  1-го порядка:

$$H_3(\delta) = H_3(0) + 2\varepsilon\delta \log \frac{1+\varepsilon}{1-\varepsilon} + O(\delta^2); \quad (14)$$

собственная информация имеет вид:

$$I\{y_{i-1} = 0, y_i = 0, y_{i+1} = 0\} = -\log \frac{(1+\varepsilon)^2}{8} + \delta \frac{1}{\ln b} \cdot \frac{2\varepsilon^2 + 4\varepsilon}{(1+\varepsilon)^2} + O(\delta^2),$$

$$\begin{aligned} I\{y_{i-1} = 1, y_i = 0, y_{i+1} = 0\} &= I\{y_{i-1} = 0, y_i = 0, y_{i+1} = 1\} = \\ &= -\log \frac{1-\varepsilon^2}{8} - \delta \frac{1}{\ln b} \cdot \frac{2\varepsilon^2}{1-\varepsilon^2} + O(\delta^2), \end{aligned}$$

$$I\{y_{i-1} = 0, y_i = 1, y_{i+1} = 0\} = -\log \frac{(1-\varepsilon)^2}{8} + \delta \frac{1}{\ln b} \cdot \frac{2\varepsilon^2 - 4\varepsilon}{(1-\varepsilon)^2} + O(\delta^2);$$

$$\begin{aligned} I\{y_{i-1} = j_1, y_i = j_2, y_{i+1} = j_3\} &= I\{y_{i-1} = 1 - j_1, y_i = 1 - j_2, y_{i+1} = 1 - j_3\}, \\ j_1, j_2, j_3 &\in \{0, 1\}. \end{aligned}$$

# Оценка энтропии 4-граммы

## Теорема

Если имеет место монобитная модель вкраплений (1)-(5), то для энтропии при  $l = 4$  справедливо асимптотическое разложение при  $\delta \rightarrow 0$  1-го порядка:

$$H_4(\delta) = H_4(0) + \frac{24\epsilon\delta}{16} \log \frac{1+\epsilon}{1-\epsilon} + O(\delta^2); \quad (15)$$

собственная информация имеет вид:

$$\begin{aligned} I\{y_{i-1} = 0, y_i = 0, y_{i+1} = 0, y_{i+2} = 0\} &= I\{y_{i-1} = 1, y_i = 1, y_{i+1} = 1, y_{i+2} = 1\} = \\ &= -\left( \log \frac{(1+\epsilon)^3}{16} + \delta \frac{-2\epsilon^3 - 8\epsilon^2 - 6\epsilon}{(1+\epsilon)^3 \ln b} \right) + O(\delta^2); \end{aligned}$$

$$\begin{aligned} I\{y_{i-1} = 0, y_i = 0, y_{i+1} = 0, y_{i+2} = 1\} &= I\{y_{i-1} = 1, y_i = 1, y_{i+1} = 1, y_{i+2} = 0\} = \\ I\{y_{i-1} = 1, y_i = 0, y_{i+1} = 0, y_{i+2} = 0\} &= I\{y_{i-1} = 0, y_i = 1, y_{i+1} = 1, y_{i+2} = 1\} = \\ &= -\left( \log \frac{(1-\epsilon)(1+\epsilon)^2}{16} + \delta \frac{2\epsilon^3 + 4\epsilon^2 - 2\epsilon}{(1-\epsilon)(1+\epsilon)^2 \ln b} \right) + O(\delta^2); \end{aligned}$$

## Оценка энтропии 4-граммы

$$\begin{aligned} I\{y_{i-1} = 0, y_i = 0, y_{i+1} = 1, y_{i+2} = 0\} &= I\{y_{i-1} = 1, y_i = 1, y_{i+1} = 0, y_{i+2} = 1\} = \\ I\{y_{i-1} = 0, y_i = 1, y_{i+1} = 0, y_{i+2} = 0\} &= I\{y_{i-1} = 1, y_i = 0, y_{i+1} = 1, y_{i+2} = 1\} = \\ &= -\left(\log \frac{(1-\varepsilon)^2(1+\varepsilon)}{16} + \delta \frac{-2\varepsilon^3 + 4\varepsilon^2 + 2\varepsilon}{(1-\varepsilon)^2(1+\varepsilon) \ln b}\right) + O(\delta^2); \\ I\{y_{i-1} = 0, y_i = 0, y_{i+1} = 1, y_{i+2} = 1\} &= I\{y_{i-1} = 1, y_i = 1, y_{i+1} = 0, y_{i+2} = 0\} = \\ &= -\left(\log \frac{(1-\varepsilon)(1+\varepsilon)^2}{16} + \delta \frac{2\varepsilon^3 - 2\varepsilon}{(1-\varepsilon)(1+\varepsilon)^2 \ln b}\right) + O(\delta^2); \\ I\{y_{i-1} = 0, y_i = 1, y_{i+1} = 1, y_{i+2} = 0\} &= I\{y_{i-1} = 1, y_i = 0, y_{i+1} = 0, y_{i+2} = 1\} = \\ &= -\left(\log \frac{(1-\varepsilon)^2(1+\varepsilon)}{16} + \delta \frac{-2\varepsilon^3 + 2\varepsilon}{(1-\varepsilon)^2(1+\varepsilon) \ln b}\right) + O(\delta^2); \\ I\{y_{i-1} = 1, y_i = 0, y_{i+1} = 1, y_{i+2} = 0\} &= I\{y_{i-1} = 0, y_i = 1, y_{i+1} = 0, y_{i+2} = 1\} = \\ &= -\left(\log \frac{(1-\varepsilon)^3}{16} + \delta \frac{2\varepsilon^3 - 8\varepsilon^2 + 6\varepsilon}{(1-\varepsilon)^3 \ln b}\right) + O(\delta^2). \end{aligned}$$

# Линейный дискриминантный анализ

Анализ последовательности  $Y = \{y_1, \dots, y_T\}$ , на основании  $(H_3(\delta), H_4(\delta))$  при фиксированном  $\varepsilon$ , тогда:

## Гипотеза

$H_0$ : последовательность  $Y$  имеет вкрапления

## Гипотеза

$H_1$ : последовательность  $Y$  не имеет вкраплений

$$\hat{\alpha} = \frac{n_0 - \nu_0}{n_0} - \text{оценка вероятности ошибки первого рода}; \quad (16)$$

$$\hat{\beta} = \frac{n_1 - \nu_1}{n_1} - \text{оценка вероятности ошибки второго рода}; \quad (17)$$

где  $n_0$  - количество заведомо пустых последовательностей,  $n_1$  - количество последовательностей с вкраплениями,  $\nu_0$  - количество верно определенных пустых последовательностей,  $\nu_1$  - количество верно определенных последовательностей с вкраплениями.

Мощность критерия:

$$\hat{w} = \frac{\nu_1}{n_1} \quad (18)$$



# Результаты ЛДА

Таблица: Результаты дискриминантного анализа при  $\varepsilon = 0.55, n = 1000$ .

$\delta$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{w}$
0.03	0.42	0.31	0.69
0.07	0.21	0.14	0.86
0.09	0.14	0.08	0.92
0.1	0.13	0.05	0.95
0.3	0.05	0.02	0.98

Таблица: Результаты дискриминантного анализа при  $\varepsilon = 0.15, n = 1000$ .

$\delta$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{w}$
0.01	0.47	0.4	0.6
0.03	0.3	0.2	0.8
0.07	0.1	0.08	0.92
0.1	0.06	0.05	0.95
0.3	0.02	0.02	0.98

# Математическая модель серий

Пусть последовательность с вкраплениями, задается следующим образом:

$$s - \nu_s, s = \overline{1, k} \quad (19)$$

где  $s$  - длина серии, а  $\nu_s$  - количество серий длины  $s$ ,  $\sum_{s=1}^k s\nu_s = T$

## Лемма

Для модели (19) асимптотическая оценка 1-го порядка вероятности вкрапления при  $\delta \rightarrow 0$  имеет вид:

$$P\{y_1 = u_1, \dots, y_T = u_T\} = \frac{1}{2^T - 1} (1 + \varepsilon)^{T - \sum_{s=1}^k \nu_s - 1 - 2} (1 - \varepsilon)^{\sum_{s=1}^k \nu_s - 2} \\ \left( (1 + \varepsilon)(1 - \varepsilon) - \delta \left( \frac{3}{4} \varepsilon^4 + \frac{5}{4} \varepsilon^3 + \frac{5}{4} \varepsilon^2 - \frac{3}{4} \varepsilon - \frac{7}{2} \right) \right) + O(\delta^2), \\ u_i \in \{0, 1\}, i = \overline{1, T}.$$

# Компьютерные эксперименты

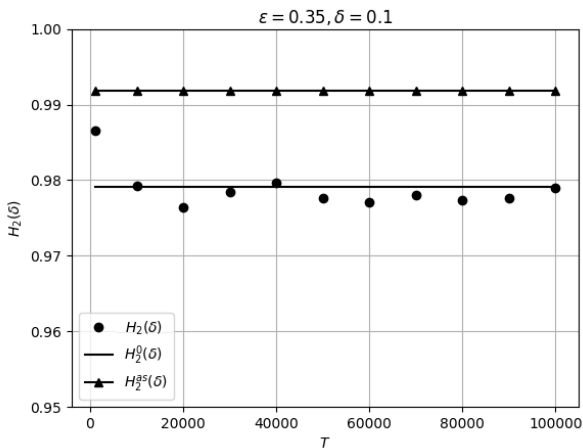


Рис.: График зависимости энтропии  $H_2(\delta)$  от длины последовательности

# Компьютерные эксперименты

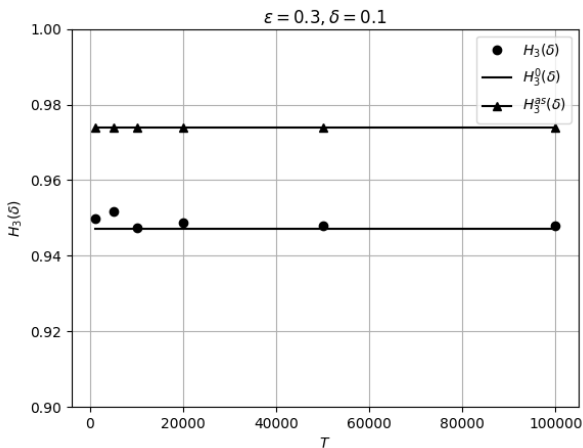


Рис.: График зависимости энтропии  $H_3(\delta)$  от длины последовательности

# Компьютерные эксперименты

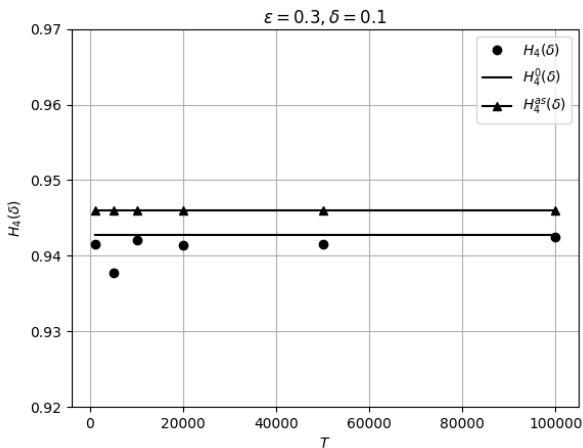


Рис.: График зависимости энтропии  $H_4(\delta)$  от длины последовательности

# Компьютерные эксперименты

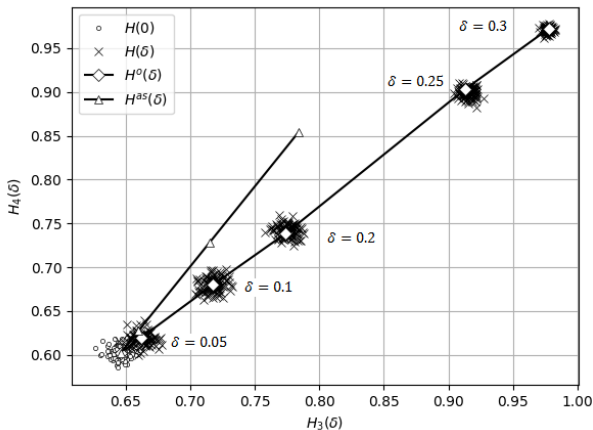









Рис.: График зависимости энтропии  $H_4(\delta)$  от  $H_3(\delta)$  при различных долях вкраплений

# Заключение

В работе получены следующие основные результаты:

1. Исследована математическая модель вкраплений в цепь Маркова 1-го порядка.
2. Получены точные значения вероятностей для всевозможных шаблонов  $l$ -граммы при  $l=2,3,4$ .
3. Получены асимптотические оценки первого порядка для энтропии  $l$ -граммы при  $l=2,3,4$ .
4. Проведен линейный дискриминантный анализ на основании асимптотических оценок для энтропии 3-граммы и 4-граммы.
5. Исследованы вероятностные свойства математической модели вкраплений в цепь Маркова 1-го порядка, задаваемой сериями.
6. Проведены компьютерные эксперименты.

# Список литературы

-  А. А. Духин: Теория информации - М.: "Гелиос АРВ", 2007.
-  А.В. Аграновский, А. В. Балакин: Стеганография, цифровые водяные знаки о стегоанализе - М.: Вузовская книга, 2009.
-  К. И. Пономарев "Параметрическая модель вкрапления и ее статистический анализ", Дискрет. матем., 21:4 (2009), 148-157.
-  Н. П. Варновский, Е. А. Голубев, О. А. Логачев: Современные направления стеганографии. Математика и безопасность информационных технологий. Материалы конференции в МГУ 28-29 октября 2004 г., МЦМНО, М., 2005, с. 32-64.
-  Ю. С. Харин [и др.]: Криптология - Минск: БГУ, 2013.
-  Ю. С. Харин, Е. В. Вечерко "Статистическое оценивание параметров модели вкраплений в двоичную цепь Маркова", Дискрет. матем., 25:2 (2013), 135-148.
-  Ю. С. Харин, Е. В. Вечерко "Распознавание вкраплений в двоичную цепь Маркова", Дискрет. матем., 27:3 (2015), 123Ц144.



Спасибо за внимание!

# ОБНАРУЖЕНИЕ ВКРАПЛЕНИЙ В ДВОИЧНУЮ ЦЕПЬ МАРКОВА НА ОСНОВЕ ЭНТРОПИЙНЫХ ХАРАКТЕРИСТИК

Андрей Чеславович Шимко

Научный руководитель: Егор Валентинович Вечерко

Минск, 2017