БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Научно-исследовательский институт
прикладных проблем математики и информатики

ТЕСТИРОВАНИЕ КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ ПРИ СЛОЖНОЙ НУЛЕВОЙ ГИПОТЕЗЕ НА ОСНОВЕ ЭНТРОПИИ ТСАЛЛИСА

Владимир Юрьевич Палуха,
Юрий Семёнович Харин

ВВЕДЕНИЕ

Генераторы случайных и псевдослучайных последовательностей являются одним из элементов систем криптографической защиты информации (СКЗИ). Стойкость СКЗИ зависит от того, насколько близка генерируемая последовательность по своим свойствам к равномерно распределённой случайной последовательности (РРСП), которая на практике называется «чисто случайной». Для проверки качества криптографических генераторов используются статистические тесты, в которых проверяется гипотеза H_0 = {{x_t} является РРСП}}. Известные наборы (батареи) статистических тестов проверяют простую нулевую гипотезу. Однако на практике возможны незначительные отклонения тестируемой последовательности от модели. Например, вероятности фрагментов длины s (s-грамм) могут отличаться от 2^{-s} на близкую к нулю величину ε. В данном докладе рассмотрено построение статистического теста на основе сложной нулевой гипотезы с применением оценки энтропии Тсаллиса.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Пусть на вероятностном пространстве (Ω, F, P) с множеством состояний $\Omega = \{\omega_1, \ldots, \omega_N\}$ определена случайная величина $x = x(\omega) = \omega$ с дискретным распределением вероятностей $p = \{p_k\}$, $p_k = P\{x = \omega_k\}$, $p_k \geq 0, \sum_{k=1}^{N} p_k = 1, k = 1, \ldots, N$. Введём в рассмотрение нулевую гипотезу H_0 = {{x_t} является РРСП} = {{x_t} – н.о.р.с.в., $p_k = p_k^0 = 1/N$, $k = 1, \ldots, N$} и альтернативу общего вида $\overline{H_0}$.

Функционал энтропии Тсаллиса:

$$S_r(p) = \frac{1}{r-1}\left(1 - \sum_{k=1}^{N} p_k^r\right), \qquad S_2(p) = 1 - \sum_{k=1}^{N} p_k^2. \qquad (1)$$

ЧАСТОТНЫЕ ОЦЕНКИ ВЕРОЯТНОСТЕЙ

Пусть имеется случайная последовательность $\{x_t : t = 1, \ldots, n\}$ объёма n из распределения вероятностей $\{p_k\}$.

$$\hat{p}_k = \frac{\nu_k}{n}, \quad \nu_k = \sum_{t=1}^{n} I\{x_t = \omega_k\} \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}, \quad I\{x_t = \omega_k\} = \begin{cases} 1, & x_t = \omega_k; \\ 0, & x_t \neq \omega_k. \end{cases} \quad (2)$$

Рассмотрим асимптотику:

$$n, N \to \infty, n/N \to \lambda, 0 < \lambda < \infty. \quad (3)$$

В асимптотике (3) для распределения вероятностей статистик $\{\nu_k\}$ справедлива аппроксимация законом Пуассона $\Pi(\lambda_k)$ с параметром $\lambda_k = np_k$. При истинной гипотезе H_0 $p_k = 1/N$, $k = 1, \ldots, N$, поэтому для всех $\{\nu_k\}$ $\lambda = n/N$. Построим с помощью $\{\nu_k\}$ оценки энтропии[1].

---

[1] Holst, L. Asymptotic normality and efficiency for certain goodness-of-fit tests / L. Holst // Biometrika. – 1972. – № 59. – P. 137–145.

СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ЭНТРОПИИ ТСАЛЛИСА

Энтропия Тсаллиса является функцией от величины

$$P_2(P) = \sum_{k=1}^{N} p_k^2 . \qquad (4)$$

Определим 2-ую нисходящую факториальную степень x:

$$x^{\underline{2}} = x(x-1). \qquad (5)$$

Несмещённая оценка для (4) основана на (5)[2]:

$$\tilde{P}_2(P) = \sum_{k=1}^{N} \frac{\nu_k^{\underline{2}}}{n^2}. \qquad (6)$$

---

[2] *Acharya, J.* Estimating Renyi Entropy of Discrete Distributions / J. Acharya, [et al.] – IEEE Transactions on Information Theory. – Vol. 63. – No. 1, 2017. – P. 38–56.

Статистическая оценка энтропии Тсаллиса, построенная с использованием оценки (6):

$$\hat{S}_2(n, N) = 1 - \sum_{k=1}^{N} \frac{\nu_k^{\underline{2}}}{n^2}. \qquad (7)$$

Теорема 1[3]. *В асимптотике* (3) *статистика* (7) *является состоятельной асимптотически несмещённой оценкой энтропии Тсаллиса и при истинной гипотезе* H_* *имеет асимптотически нормальное распределение с параметрами*:

$$\mu_{S,2} = 1 - \frac{1}{N}, \qquad (8)$$

$$\sigma_{S,2}^2 = \frac{2}{Nn^2}. \qquad (9)$$

---

[3] Статистические оценки энтропии Реньи и Тсаллиса и их использование для проверки гипотез о «чистой случайности» / Ю. С. Харин, В. Ю. Палуха // Весці НАН Беларусі. Серыя фізіка-матэматычных навук. – 2016. – № 2. – С. 37–47.

СЛОЖНАЯ ГИПОТЕЗА

Введём в рассмотрение сложную нулевую гипотезу $H_0^{(\varepsilon)}$, согласно которой для распределения вероятностей $\{p_k\}$ справедливо

$$p_k = \frac{1}{N} + \varepsilon_k,\ \varepsilon_k \geq -\frac{1}{N},\ \varepsilon_k = O\left(\frac{1}{N}\right),\ k = 1, \ldots, N,$$
$$\sum_{k=1}^{N} \varepsilon_k = 0,\ \sum_{k=1}^{N} \varepsilon_k^2 = \varepsilon^2,\ 0 < \varepsilon \leq \varepsilon_+. \qquad (10)$$

Заметим, что при $\varepsilon = 0$ получаем гипотезу H_0, поскольку $\varepsilon_k = 0, k = 1, \ldots, N$.

В асимптотике (3) для распределения вероятностей статистик $\{\nu_k\}$ справедлива аппроксимация законом Пуассона $\Pi(\lambda_k)$ с параметром

$$\lambda_k = np_k = \frac{n}{N} + n\varepsilon_k = \lambda + n\varepsilon_k. \qquad (11)$$

Теорема 2[4]. *В асимптотике* (3) *статистика* (7) *является состоятельной асимптотически несмещённой оценкой энтропии Тсаллиса и при истинной гипотезе* $H_0^{(\varepsilon)}$ *имеет асимптотически нормальное распределение с параметрами:*

$$\mu_{S,2}^{(\varepsilon)} = 1 - \frac{1}{N} - \sum_{k=1}^{N} \varepsilon_k^2 = \mu_{S,2} - \varepsilon^2, \tag{12}$$

$$\begin{aligned} \sigma_{S,2}^{2(\varepsilon)} &= \frac{2}{Nn^2} + \frac{2}{n^2}\left(2\frac{n}{N}+1\right)\sum_{k=1}^{N}\varepsilon_k^2 + \frac{4}{n}\left(\sum_{k=1}^{N}\varepsilon_k^3 - \left(\sum_{k=1}^{N}\varepsilon_k^2\right)^2\right) = \\ &= \sigma_{S,2}^2 + \frac{2}{n^2}\left(2\lambda+1\right)\varepsilon^2 + \frac{4}{n}\left(\sum_{k=1}^{N}\varepsilon_k^3 - \varepsilon^4\right). \end{aligned} \tag{13}$$

[4] Палуха, В. Ю. Статистическое тестирование криптографических генераторов на основе сложной нулевой гипотезы / В. Ю. Палуха, Ю. С. Харин // Теоретическая и прикладная криптография: материалы II Международной научной конференции, Минск, 19–20 октября 2023 г. – Минск: БГУ, 2023. – С. 185–193.

Лемма[4]. *Для* (12) и (13) *с учётом* $\varepsilon \le \varepsilon_+$ *при* $0 < \varepsilon_+ \le \frac{1}{\sqrt{2}}$ *справедливы оценки:*

$$\mu_{S,2}^{(\varepsilon)} = \mu_{S,2} - \varepsilon^2 \ge \mu_{S,2}^{(\varepsilon_+)} = \mu_{S,2} - \varepsilon_+^2, \tag{15}$$

$$\sigma_{S,2}^{2(\varepsilon)} < \sigma_{S,2}^{2(\varepsilon_+)} = \sigma_{S,2}^2 + \frac{2}{n^2}\left(2\lambda + 1\right)\varepsilon_+^2 + \frac{4\varepsilon_+^3}{n}\left(1 - \varepsilon_+\right). \tag{16}$$

Теорема 3[4]. *Пусть* $\alpha \in (0,1)$ – *заданный уровень значимости. Тогда в асимптотике* (3) *при истинной гипотезе* $H_0^{(\varepsilon)}$ *и* $0 < \varepsilon_+ \le \frac{1}{\sqrt{2}}$ *для статистики* (7) *справедливо*

$$P\left\{\Delta_- < \hat{S}_2 < \Delta_+\right\} \ge 1 - \alpha,$$

$$\Delta_- = \mu_{S,2}^{(\varepsilon_+)} - \sigma_{S,2}^{(\varepsilon_+)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \quad \Delta_+ = \mu_{S,2} + \sigma_{S,2}^{(\varepsilon_+)}\Phi^{-1}\left(1 - \frac{\alpha}{2}\right). \tag{17}$$

РЕШАЮЩЕЕ ПРАВИЛО

На основе (17) построим решающее правило. Пусть вычислена статистика $\hat{S}_2$ и задан уровень значимости $\alpha \in (0,1)$, тогда решающее правило имеет вид:

$$\text{принимается} \begin{cases} H_0^{(\varepsilon)}, \text{ если } \Delta_- < \hat{S}_2 < \Delta_+, \\ \overline{H_0^{(\varepsilon)}}, \text{ иначе,} \end{cases} \tag{18}$$

$$\Delta_- = \mu_{S,2}^{(\varepsilon_+)} - \sigma_{S,2}^{(\varepsilon_+)} \Phi^{-1}\left(1-\frac{\alpha}{2}\right), \quad \Delta_+ = \mu_{S,2} + \sigma_{S,2}^{(\varepsilon_+)} \Phi^{-1}\left(1-\frac{\alpha}{2}\right), \quad \varepsilon_+ \le \frac{1}{\sqrt{2}}.$$

где $\Phi(\cdot)$ – функция распределения стандартного нормального закона[5].

[5] *Харин, Ю. С.* Теория вероятностей, математическая и прикладная статистика / Ю. С. Харин, Н. М. Зуев, Е. Е. Жук. – Минск: БГУ, 2011. – 463 с.

КОМПЬЮТЕРНЫЕ ЭКСПЕРИМЕНТЫ

Проведено две серии компьютерных экспериментов. В каждой из них были сгенерированы по 20 последовательностей одинаковой длины из дискретного распределения (10), при этом вектор вероятностей p выбирался случайно равновероятно на гиперсфере с центром в равномерном распределении и заданным радиусом ε.

В первой серии экспериментов длина последовательностей составляла 1 МБ, мощность алфавита равнялась N = 256, радиус гиперсферы равнялся $\varepsilon = 2^{-12}$. Для каждой последовательности вычислялась оценка энтропии Тсаллиса (3) при N = 256. Для двух последовательностей простая гипотеза H_0 на уровне значимости $\alpha = 0{,}05$ была отклонена (значения оценок оказались меньше нижнего порога), однако сложная гипотеза $H_0^{(\varepsilon)}$ при $\varepsilon_+ = 2^{-12}$ была принята. Численные значения приведены в таблице.

$\mu_{S,2} = 0{,}99609375$	$\Delta_- = 0{,}996093584787389$
$\mu_{S,2}^{(\varepsilon_+)} = 0{,}996093690395355$	$\Delta_-^{(\varepsilon_+)} = 0{,}996093514552483$
$\hat{S}_2\ (1) = 0{,}996093579678927$	$\hat{S}_2\ (2) = 0{,}99609356833389$

Во второй серии экспериментов длина последовательностей составляла 32 МБ, мощность алфавита равнялась $N = 65536$, радиус гиперсферы равнялся $\varepsilon = 2^{-16}$. Для каждой последовательности вычислялась оценка энтропии Тсаллиса (3) при $N = 65536$. Для одной последовательности простая гипотеза H_0 на уровне значимости $\alpha = 0{,}05$ была отклонена (значение оценки оказалось меньше нижнего порога), однако сложная гипотеза $H_0^{(\varepsilon)}$ при $\varepsilon_+ = 2^{-16}$ была принята. Ещё для одной последовательности была отклонена и сложная гипотеза. Численные значения приведены в таблице.

$\mu_{S,2} = 0{,}99998474121093 7$	$\Delta_- = 0{,}99998474056557 5$
$\mu_{S,2}^{(\varepsilon_+)} = 0{,}99998474017606 7$	$\Delta_-^{(\varepsilon_+)} = 0{,}99998474033022 9$
$\hat{S}_2\ (1) = 0{,}99998474039958 9$	$\hat{S}_2\ (2) = 0{,}99998474017606 7$

СПАСИБО ЗА ВНИМАНИЕ!