БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Научно-исследовательский институт
прикладных проблем математики и информатики

ТЕСТИРОВАНИЕ КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ СЛУЧАЙНЫХ И ПСЕВДОСЛУЧАЙНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ НА ОСНОВЕ ЭНТРОПИЙНЫХ ПРОФИЛЕЙ

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)[1], которая на практике называется «чисто случайной». Для проверки качества криптографических генераторов используются статистические тесты, в которых проверяется гипотеза $H_* = \{\{x_t\}$ является РРСП$\}\}$. В качестве тестовой статистики целесообразно использовать статистические оценки энтропии[2].

---

[1] Криптология / Ю. С. Харин [и др.]. – Минск: БГУ, 2013. – 512 с.

[2] Харин, Ю. С. Энтропийный анализ криптографических генераторов случайных и псевдослучайных последовательностей / Ю. С. Харин, В. Ю. Палуха // Веснік сувязі. – 2017. – № 6 (146). – С. 40–43.

ФУНКЦИОНАЛЫ ЭНТРОПИИ

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 $p = \{p_k\},\ p_k = P\{x = \omega_k\},\ p_k \geq 0, \sum_{k=1}^{N} p_k = 1, k = 1, \ldots, N$.

Функционалы энтропии:

Энтропия Шеннона	$H(p) = -\sum_{i=1}^{N} p_i \ln p_i$
Энтропия Реньи	$H_r(p) = \frac{1}{1-r} \ln\left(\sum_{i=1}^{N} p_i^r\right),\ r \in \mathbb{N}, r > 1.$
Энтропия Тсаллиса	$S_r(p) = \frac{1}{r-1}\left(1 - \sum_{i=1}^{N} p_i^r\right),\ r \in \mathbb{N}, r > 1.$

ЧАСТОТНЫЕ ОЦЕНКИ ВЕРОЯТНОСТЕЙ

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

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

Рассмотрим асимптотику:

$$n, N \to \infty, n/N \to \lambda, 0 < \lambda < \infty. \quad (2)$$

В асимптотике (2) для распределения вероятностей статистик $\{\nu_k\}$ справедлива аппроксимация законом Пуассона $\Pi(\lambda_k)$ с параметром $\lambda_k = np_k$. При истинной гипотезе H_* $p_k = 1/N$, $k = 1, \ldots, N$, поэтому для всех $\{\nu_k\}$ $\lambda = n/N$. Построим с помощью $\{\nu_k\}$ оценки энтропии[3].

[3] *Палуха, В. Ю.* Статистические тесты на основе оценок энтропии для проверки гипотез о равномерном распределении случайной последовательности / В. Ю. Палуха // Весці НАН Беларусі. Серыя фізіка-матэматычных навук. – 2017. – № 1. – С. 79–88.

СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ЭНТРОПИИ ШЕННОНА

Оценка энтропии Шеннона на основе частотных статистик (1):

$$\hat{H} = \hat{H}(n, N) = -\sum_{k=1}^{N} \hat{p}_k \ln \hat{p}_k = -\sum_{k=1}^{N} \frac{\nu_k}{n} \ln \frac{\nu_k}{n} = \ln n - \frac{1}{n} \sum_{k=1}^{N} \nu_k \ln \nu_k . \qquad (3)$$

Теорема 1^3. *В асимптотике* (2) *статистика* (3) *при гипотезе* H_* *имеет асимптотически нормальное распределение с параметрами*

$$\mu_H = \ln n - e^{-\lambda} \sum_{k=1}^{+\infty} \frac{\ln(k+1)\lambda^k}{k!}, \qquad (4)$$

$$\sigma_H^2 = \frac{e^{-\lambda}}{n} \sum_{k=1}^{+\infty} \frac{(k+1)\lambda^k}{k!} \ln^2(k+1) - \frac{e^{-2\lambda}}{N} \left(\sum_{k=1}^{+\infty} \frac{\ln(k+1)\lambda^k}{k!} \right)^2 - \qquad (5)$$

$$- \frac{e^{-2\lambda}}{n} \left(\sum_{k=1}^{+\infty} \ln(k+1) \frac{\lambda^k}{k!} (k+1-\lambda) \right)^2 .$$

СТАТИСТИЧЕСКОЕ ОЦЕНИВАНИЕ ЭНТРОПИИ РЕНЬИ И ТСАЛЛИСА

Энтропии Реньи и Тсаллиса являются функциями от величины

$$P_r(P) = \sum_{k=1}^{N} p_k^r . \qquad (6)$$

Определим r-ую нисходящую факториальную степень x:

$$x^{\underline{r}} = x(x-1)\ldots(x-r+1) = \frac{x!}{(x-r)!} = \sum_{i=0}^{r} s(r,i)x^i , \qquad (7)$$

где $s(r, i)$ – число Стирлинга первого рода; при $x < r$ полагают $x^{\underline{r}} ::= 0$.

Несмещённая оценка для (6) основана на (7)[4]:

---

[4] *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.

$$\tilde{P}_r(P) = \sum_{k=1}^{N} \frac{\nu_k^{\underline{r}}}{n^r}. \qquad (8)$$

Полагая, что случайная величина ν имеет распределение Пуассона с параметром λ, т.е. $\mathcal{L}\{\nu\} = \Pi(\lambda)$, получим согласно[2] $E\{\nu^{\underline{r}}\} = \lambda^r$. Кроме того, согласно [5], $E\{\nu^r\} = \sum_{i=0}^{r} S(r,i)\lambda^i$, где $S(r,\ i)$ – число Стирлинга второго рода.

Статистические оценки энтропии Реньи и Тсаллиса, построенные с использованием оценки (8):

$$\hat{H}_r(n, N) = \frac{1}{1-r} \ln\left(\sum_{k=1}^{N} \hat{p}_k^r \right) = \ln n + \frac{1}{r-1}\left(\ln n - \ln \sum_{k=1}^{N} \nu_k^{\underline{r}} \right), \qquad (9)$$

---

[5] *Riordan, J.* (1937). Moment recurrence relations for binomial, Poisson and hypergeometric frequency distributions / J. Riordan // Annals of Mathematical Statistics. – 1937. – Vol. 8, № 2. – P. 103–111.

$$\hat{S}_r(n, N) = \frac{1}{r-1}\left(1 - \sum_{k=1}^{N} \hat{p}_k^r\right) = \frac{1}{r-1}\left(1 - \frac{1}{n^r}\sum_{k=1}^{N} \nu_k^r\right). \qquad (10)$$

Теорема 2^3. *В асимптотике* (2) *статистика* (10) *является состоятельной асимптотически несмещённой оценкой энтропии Тсаллиса и при истинной гипотезе* H_* *имеет асимптотически нормальное распределение с параметрами*:

$$\mu_{S,r} = \frac{1}{r-1}\left(1 - \frac{1}{N^{r-1}}\right), \qquad (11)$$

$$\sigma^2_{S,r} = \frac{\lambda^{r-1}}{(r-1)^2 n^{2r-1}}\left(\sum_{i=2}^{r} s(r,i)\sum_{j=1}^{i-1} C_i^j r^{i-j}\sum_{k=1}^{j} S(j,k)\lambda^k - r^2\lambda^{r-1} + r!\right). \qquad (12)$$

Следствие 1. *При* $r = 2$ *для математического ожидания и дисперсии асимптотического распределения оценки* (10) *справедливы выражения*:

$$\mu_{S,2} = 1 - \frac{1}{N},$$

$$\sigma^2_{S,2} = \frac{2}{Nn^2}.$$

Теорема 3^3. *В асимптотике* (2) *статистика* (9) *является состоятельной оценкой энтропии Реньи и при истинной гипотезе* H_* *имеет асимптотически нормальное распределение с параметрами*:

$$\mu_{H,r} = \ln N, \tag{13}$$

$$\sigma^2_{H,r} = \frac{\sum_{i=2}^{r} s(r,i) \sum_{j=1}^{i-1} C_i^j r^{i-j} \sum_{k=1}^{j} S(j,k) \lambda^k - r^2 \lambda^{r-1} + r!}{(r-1)^2 n \lambda^{r-1}}. \tag{14}$$

Следствие 2. *При* $r = 2$ *для дисперсии асимптотического распределения вероятностей оценки* (9) *справедливо выражение*:

$$\sigma^2_{H,2} = \frac{2}{n\lambda}.$$

РЕШАЮЩЕЕ ПРАВИЛО

Пусть $\alpha \in (0,1)$ – заданный уровень значимости. Введём обозначения: $\hat{h}$ – статистическая оценка энтропии Шеннона (3), Реньи (9) или Тсаллиса (10), μ_h – асимптотическое математическое ожидание статистической оценки энтропии Шеннона (4), Реньи (13) или Тсаллиса (11), σ_h^2 – асимптотическая дисперсия статистической оценки энтропии Шеннона (5), Реньи (14) или Тсаллиса (12) при истинной гипотезе H_*. Вычислим для наблюдаемой последовательности статистику $\hat{h}$. Решающее правило, основанное на статистике $\hat{h}$, имеет вид:

$$\begin{cases} H_*, & \text{если } t_- < \hat{h} < t_+; \\ \overline{H_*}, & \text{в противном случае,} \end{cases} \qquad t_\pm = \mu_h \pm \sigma_h \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \qquad (15)$$

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

ЭНТРОПИЙНЫЙ ПРОФИЛЬ

Пусть генератор порождает двоичную выходную последовательность $\{y_\tau\}$, $\tau = 1, \ldots, T$. «Нарежем» её на непересекающиеся подряд идущие фрагменты длины s (s-граммы): $X^{(t)} = (X_j^{(t)}) = (y_{(t-1)s+1}, \ldots, y_{ts}) \in \{0,1\}^s$, $t = 1, \ldots, n = [T / s]$. Из полученных s-грамм сформируем новую последовательность $\{x_t\}$ из алфавита мощности $N = 2^s$ по правилу $x_t = \sum_{j=1}^{s} 2^{j-1} X_j^{(t)} + 1$.

На основе критерия (15) мы можем вычислить последовательности нормированных отклонений оценки энтропии от математического ожидания в зависимости от s, которые назовём **энтропийными профилями**:

$$\chi(s) = \frac{\hat{h}(s) - \mu_h(s)}{\sigma_h(s)\Phi^{-1}(1-\alpha/2)}, \; s = 1, \ldots, s_+. \qquad (16)$$

Тестирование с помощью профиля позволяет выносить решение о принятии или отклонении гипотезы H_* на основе решающего правила (15) по последовательности значений $\chi(s)$ для различных s; такое решение видится более аргументированным, чем при принятии его по результатам применения теста (15) для одного значения s.

Представляют теоретический и практический интерес задачи исследования стохастической зависимости статистик $\chi(s)$ в (16). Исследуем зависимость соседних статистик $\chi(s)$ и $\chi(s+1)$. Для этого вначале исследуем зависимость частотных оценок (1). Пусть $\{\hat{p}_i(s)\}$ и $\{\hat{p}_k(s+1)\}$ – частотные оценки (1), вычисленные для s- и $(s+1)$-грамм соответственно, $i = 0, \ldots, 2^s - 1, k = 0, \ldots, 2^{s+1} - 1.$ Справедлива следующая теорема о коэффициенте корреляции частотных оценок вероятностей s- и $(s+1)$-грамм.

Теорема 4. *При истинной гипотезе* H_* *для коэффициента корреляции*

$$Corr_*\left\{\hat{p}_i(s),\ \hat{p}_k(s+1)\right\} = \frac{\mathrm{cov}_*\left\{\hat{p}_i(s),\ \hat{p}_k(s+1)\right\}}{\sqrt{D_*\left\{\hat{p}_i(s)\right\} D_*\left\{\hat{p}_k(s+1)\right\}}}$$

частотных оценок вероятностей произвольной пары s- *и* $(s+1)$-*грамм* $(i,k),\ i=0,\ldots,2^s-1,\ k=0,\ldots,2^{s+1}-1,$ *справедлива двусторонняя оценка*

$$-2\sqrt{\frac{s}{(s+1)\left(2^s-1\right)\left(2^{s+1}-1\right)}} \le$$

$$\le Corr_*\left\{\hat{p}_i(s),\ \hat{p}_k(s+1)\right\} \le \frac{2^{s+2}-2s-4}{\sqrt{s(s+1)\left(2^s-1\right)\left(2^{s+1}-1\right)}}.$$

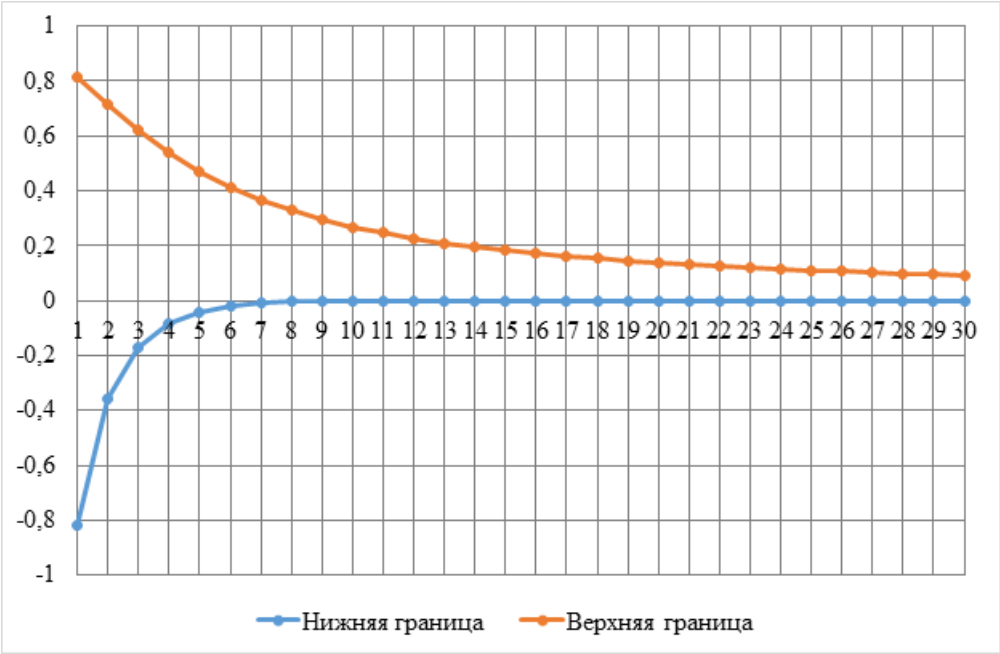

Рисунок 1. – Границы коэффициента корреляции частотных оценок

На рисунке 1 представлена зависимость верхней и нижней границы коэффициента корреляции частотных оценок вероятностей s- и $(s + 1)$-грамм от s. Из рисунка видно, что модуль коэффициента корреляции стремится к 0 с ростом s. Следовательно, зависимость между оценками энтропии h_s и h_{s+1} по s- и $(s + 1)$-граммам также будет слабой, что позволяет выносить решение о качестве генератора по его энтропийному профилю, пренебрегая этой зависимостью.

КОМПЬЮТЕРНЫЕ ЭКСПЕРИМЕНТЫ

На рисунке 2 представлен энтропийный профиль Тсаллиса выходной последовательности программного самосжимающего генератора на основе РСЛОС с многочленом $x^{24} + x^{11} + x^5 + x^2 + 1$ при фиксированном значении $\lambda = 2$ при $r = 2$ на уровне значимости $\alpha = 0.05$.

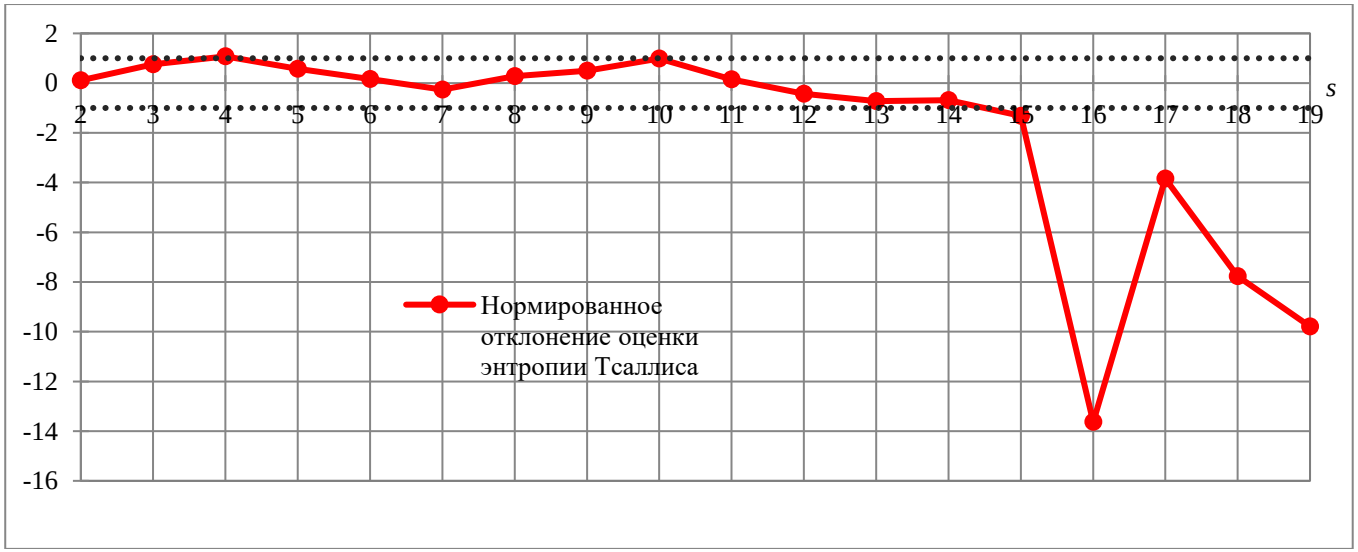

Рисунок 2. – Энтропийный профиль Тсаллиса самосжимающего генератора

ПРОГРАММНЫЙ КОМПЛЕКС

В настоящее время в НИИ ППМИ ведётся разработка программного комплекса (ПК) энтропийного анализа выходных последовательностей криптографических генераторов, который позволит автоматизировать процесс принятия решений и визуализации энтропийных профилей последовательностей. На рисунке 3 представлен результат работы ПК по энтропийному анализу выходной двоичной последовательности физического генератора[6] длиной $T = 125 \cdot 2^{25}$ бит. Выведены на экран энтропийные профили Шеннона, Реньи и Тсаллиса.

[6] speedtest-500MB.bin [Electronic resource] // Humboldt Berlin University, Faculty of Mathematics and Natural Sciences, Department of Physics. – Mode of access: http://qrng.physik.hu-berlin.de/files/speedtest-500MB.bin.

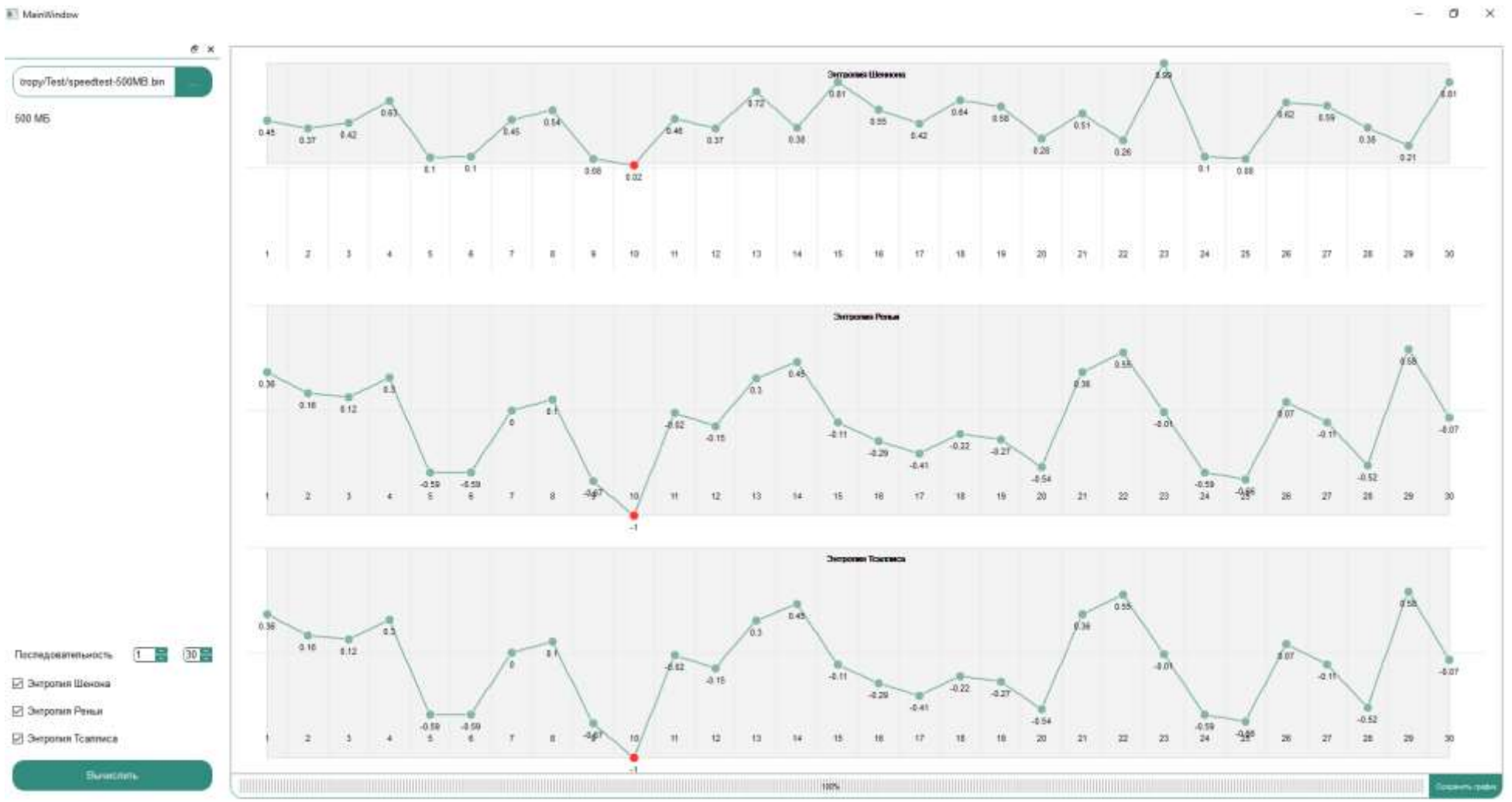

Рисунок 3. – Энтропийные профили физического генератора

СПАСИБО ЗА ВНИМАНИЕ!