БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
Научно-исследовательский институт
прикладных проблем математики и информатики

ТЕСТИРОВАНИЕ ПСЕВДОСЛУЧАЙНЫХ КРИПТОГРАФИЧЕСКИХ ГЕНЕРАТОРОВ НА ОСНОВЕ ЭНТРОПИЙНЫХ СТАТИСТИК ТСАЛЛИСА

В.Ю. Палуха, Ю.С. Харин
Минск, Беларусь
palukha@bsu.by, kharin@bsu.by

ВВЕДЕНИЕ

Для проверки качества генераторов случайных и псевдослучайных последовательностей используются статистические тесты. В данном докладе рассматриваются тесты на основе статистических оценок энтропии Тсаллиса для проверки гипотезы $H_* = \{\{x_t\}$ является РРСП$\}\}$ о том, что наблюдаемая последовательность является равномерно распределённой случайной последовательностью.

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 $P = \{p_k\}$, $p_k = P\{x = \omega_k\}$, $p_k \geq 0$, $\sum_{k=1}^{N} p_k = 1$, $k = 1, \ldots, N$. Гипотеза H_* принимает вид $\{\{x_t\}$ – н.о.р.с.в., $p_k = p_k^0 = 1/N$, $k = 1, \ldots, N\}$.

ОЦЕНКА ЭНТРОПИИ ТСАЛЛИСА

Энтропия Тсаллиса определяется формулой

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

Пусть имеется случайная выборка $X_n = \{x_t : t = 1, \ldots, n\}$ объёма n из распределения вероятностей $\{p_k\}$. Построим частотные оценки распределения вероятностей $\{p_k : k = 1, \ldots, N\}$ с использованием факториальных степеней:

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

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

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

Статистическая оценка энтропии Тсаллиса (1) $\hat{S}_r(n, N)$, построенная по подстановочному принципу с использованием частотных оценок вероятностей (2), записывается следующим образом:

$$\hat{S}_r(n, N) = \frac{1}{r-1}\left(1 - \sum_{k=1}^{N} \frac{\nu_k^r}{n^r}\right). \quad (3)$$

Рассмотрим асимптотику[1], в которой длительность наблюдения n и число значений N растут синхронно:

$$n, N \to \infty, n / N \to \lambda, 0 < \lambda < \infty. \quad (4)$$

Справедлива теорема[2] об асимптотическом распределении вероятностей статистики (3).

---

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

[2] 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.

Теорема. В асимптотике (4) статистика (3) является состоятельной асимптотически несмещённой оценкой энтропии Тсаллиса и при истинной гипотезе H_* имеет асимптотически нормальное распределение $\mathcal{L}\left\{\dfrac{\hat{H}-\mu_H}{\sigma_H}\right\} \to \mathcal{N}_1(0,1)$:

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

$$\sigma^2_{S,r} = \frac{\lambda^{r-1}}{(r-1)^2 n^{2r-1}}\left(\sum_{i=1}^{r} s(r,i)\sum_{j=1}^{i-1} C_i^j r^{i-j}\sum_{k=1}^{j} S(j,k)\lambda^k - r^2\lambda^{r-1} + r!\right), \quad (5)$$

где $S(r, i)$ – число Стирлинга второго рода.

Следствие. При $r = 2$ для математического ожидания и дисперсии асимптотического распределения оценки (4) справедливы выражения:

$$\mu_{S,2} = 1-\frac{1}{N}, \sigma^2_{S,2} = \frac{2}{Nn^2}. \quad (5)$$

РЕШАЮЩЕЕ ПРАВИЛО

Пусть $\varepsilon \in (0,1)$ – заданный уровень значимости. Знание асимптотического распределения точечной оценки (3) позволяет построить интервальную оценку энтропии Тсаллиса:

$$\text{с вероятностью } 1-\varepsilon \; S_r(P) \in (S_-, S_+), \; S_\pm = \mu_{S,r} \pm \sigma_{S,r} \Phi^{-1}\left(1-\frac{\varepsilon}{2}\right), \quad (6)$$

где $\Phi(\cdot)$ – функция распределения стандартного нормального закона. Решающее правило, основанное на интервальной оценке (6), имеет вид:

$$\text{принимается} \begin{cases} H_*, & \text{если } t_- < \hat{H}(n,N) < t_+; \\ \overline{H_*}, & \text{в противном случае,} \end{cases} \qquad t_\pm = \mu_H \pm \sigma_H \Phi^{-1}\left(1-\frac{\varepsilon}{2}\right). \quad (7)$$

ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

Рассматривались последовательности нелинейного регистра сдвига (G_1) с функцией обратной связи 24-го порядка $x_0 \oplus x_1 \oplus x_8 \oplus x_9 \oplus x_{15} \oplus x_7 x_{18}$, прореживающего генератора (G_2) с порождающим РСЛОС с многочленом $x^{13} + x^8 + x^5 + x^3 + 1$ и управляющим РСЛОС с многочленом $x^{11} + x^2 + 1$, самосжимающего генератора (G_3) на основе РСЛОС с многочленом $x^{24} + x^{11} + x^5 + x^2 + 1$, последовательность алгоритма ГОСТ 28147-89, работающего в режиме гаммирования (G_4), с нулевыми ключом и синхропосылкой. Все выходные последовательности «разрезаются» 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 формируется новая последовательность $\{x_t\}$ из алфавита мощности $N = 2^s$ по правилу $x_t = \sum_{j=1}^{s} 2^{j-1} X_j^{(t)} + 1$.

Первая серия экспериментов проводилась на выходных последовательностях фиксированной длины T = 1Мбайт. Таким образом, при фиксированном T и с ростом s меняется значение λ. На рисунках 1–4 для указанных выше четырёх псевдослучайных генераторов представлены графики зависимости нормированных отклонений оценки энтропии Тсаллиса (3) от асимптотического математического ожидания (5): $\dfrac{\hat{S}_2 - \mu_{S,2}}{\sigma_{S,2}\Phi^{-1}(1-\varepsilon/2)}$, на уровне значимости $\varepsilon = 0.05$ в зависимости от s. Зависимость λ от s приведена на рисунке 5. Как видно из рисунков, статистические оценки энтропии Тсаллиса генераторов G_1, G_2, G_3 выходят за границы интервала (–1; 1), что означает отклонение гипотезы H_*. Возвращение внутрь доверительного интервала при больших значениях порядков s объясняется недостаточной длиной последовательности T. Что касается генератора G_4, то выходы за границы доверительного интервала незначительны и позволяют принять гипотезу H_* при $s \leq 23$.

Рисунок 1 – Нормированное отклонение оценки энтропии нелинейного регистра сдвига

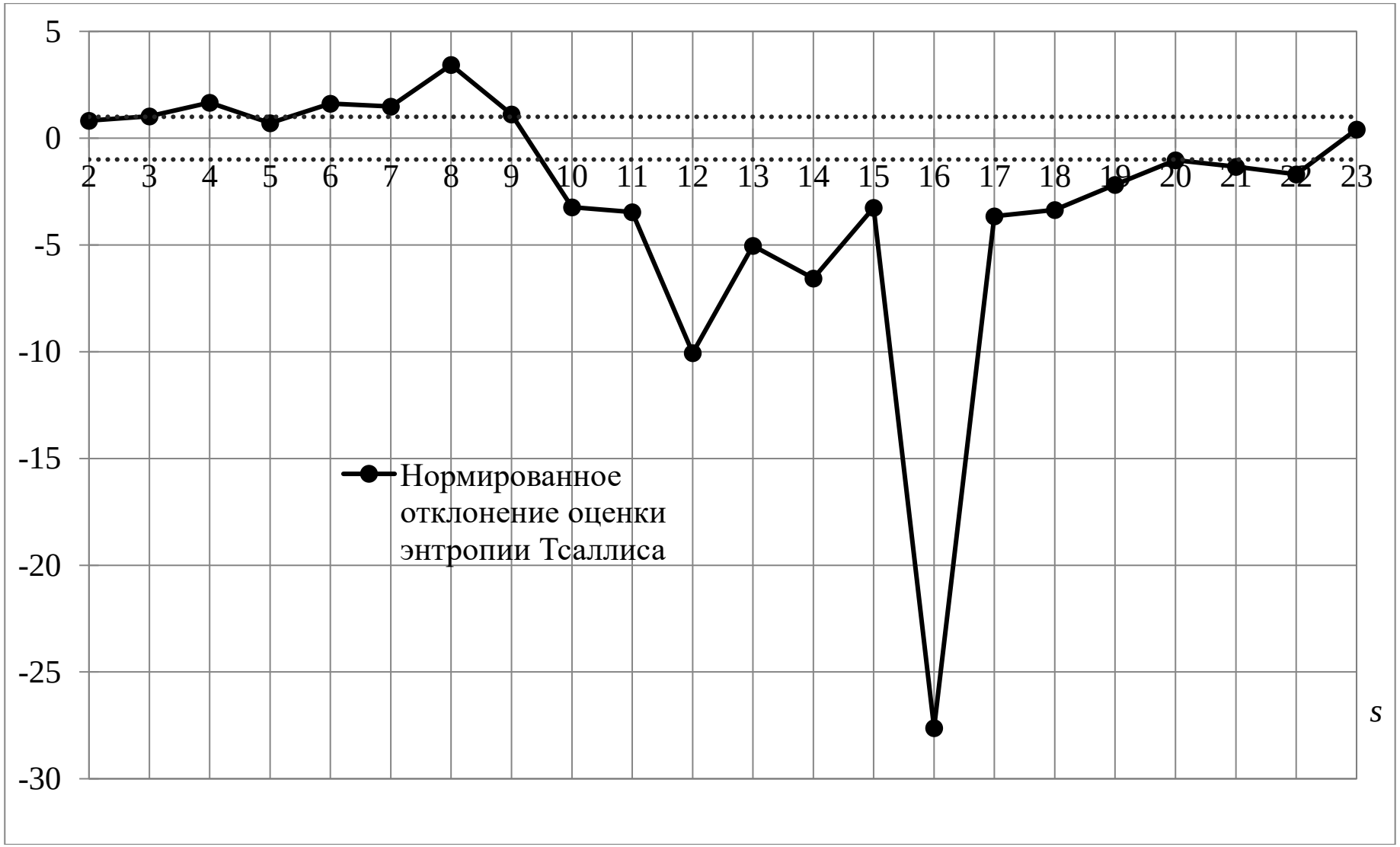

Рисунок 2 – Нормированное отклонение оценки энтропии прореживающего генератора

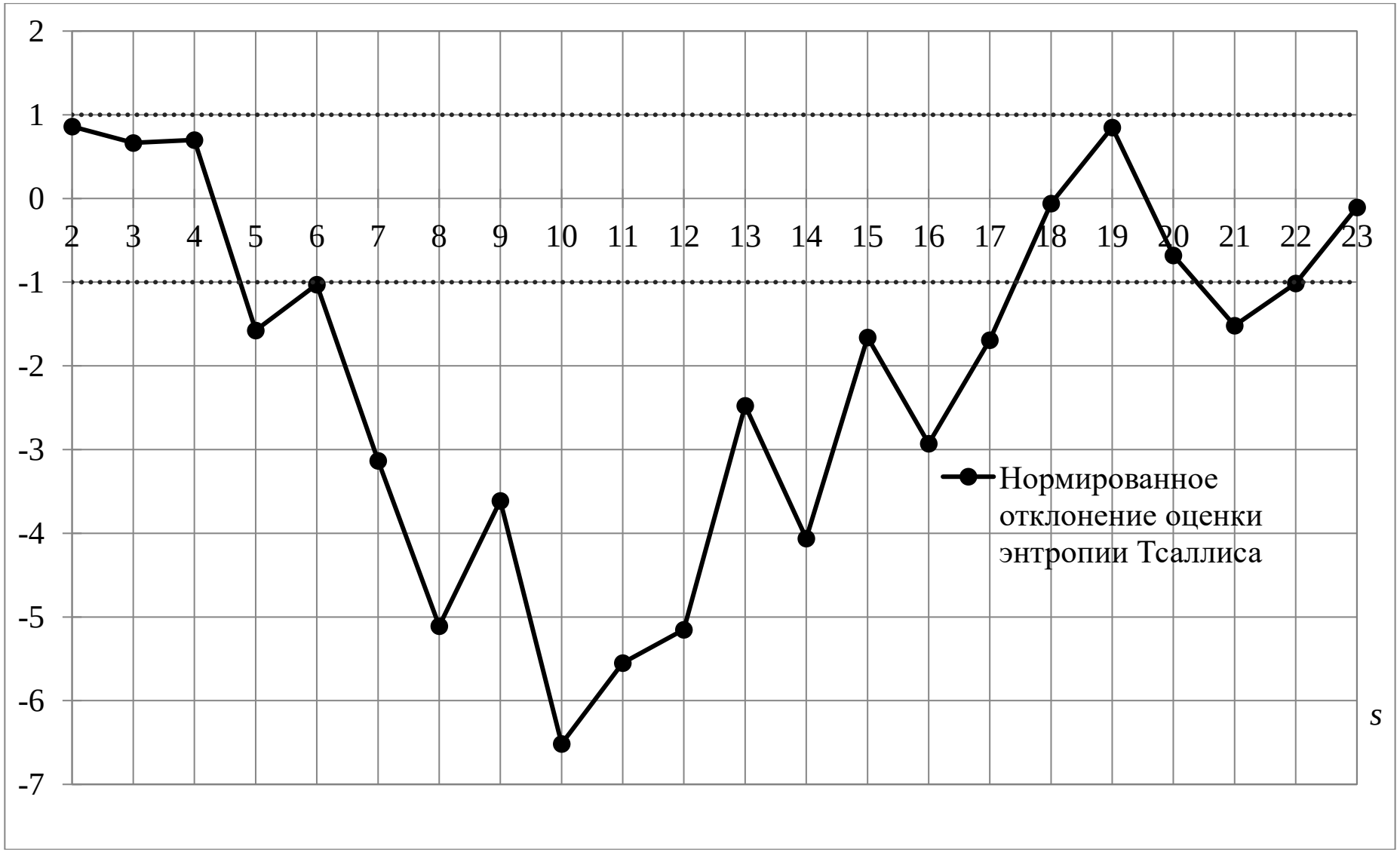

Рисунок 3 – Нормированное отклонение оценки энтропии самосжимающего генератора

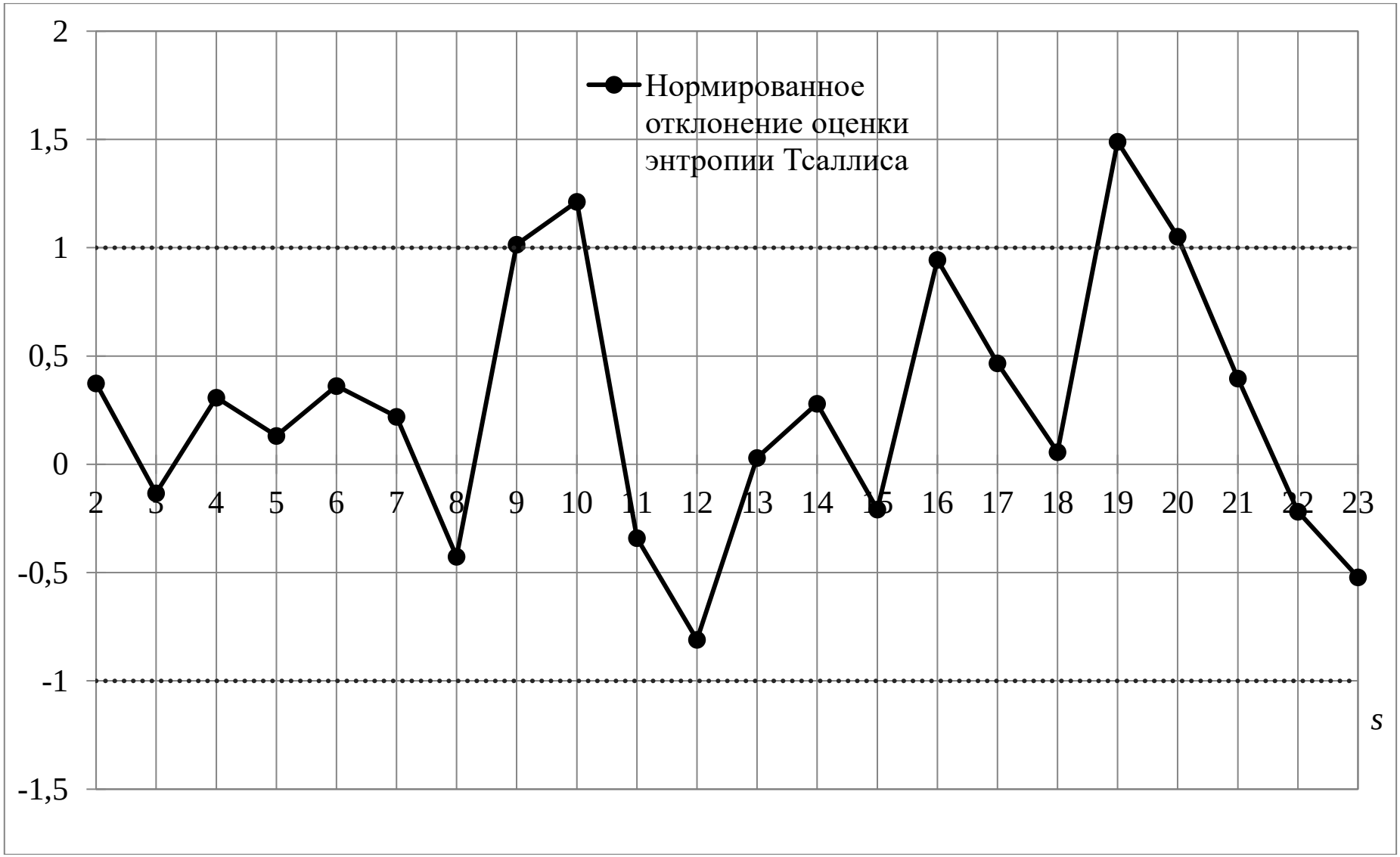

Рисунок 4 – Нормированное отклонение оценки энтропии алгоритма ГОСТ 28147-89

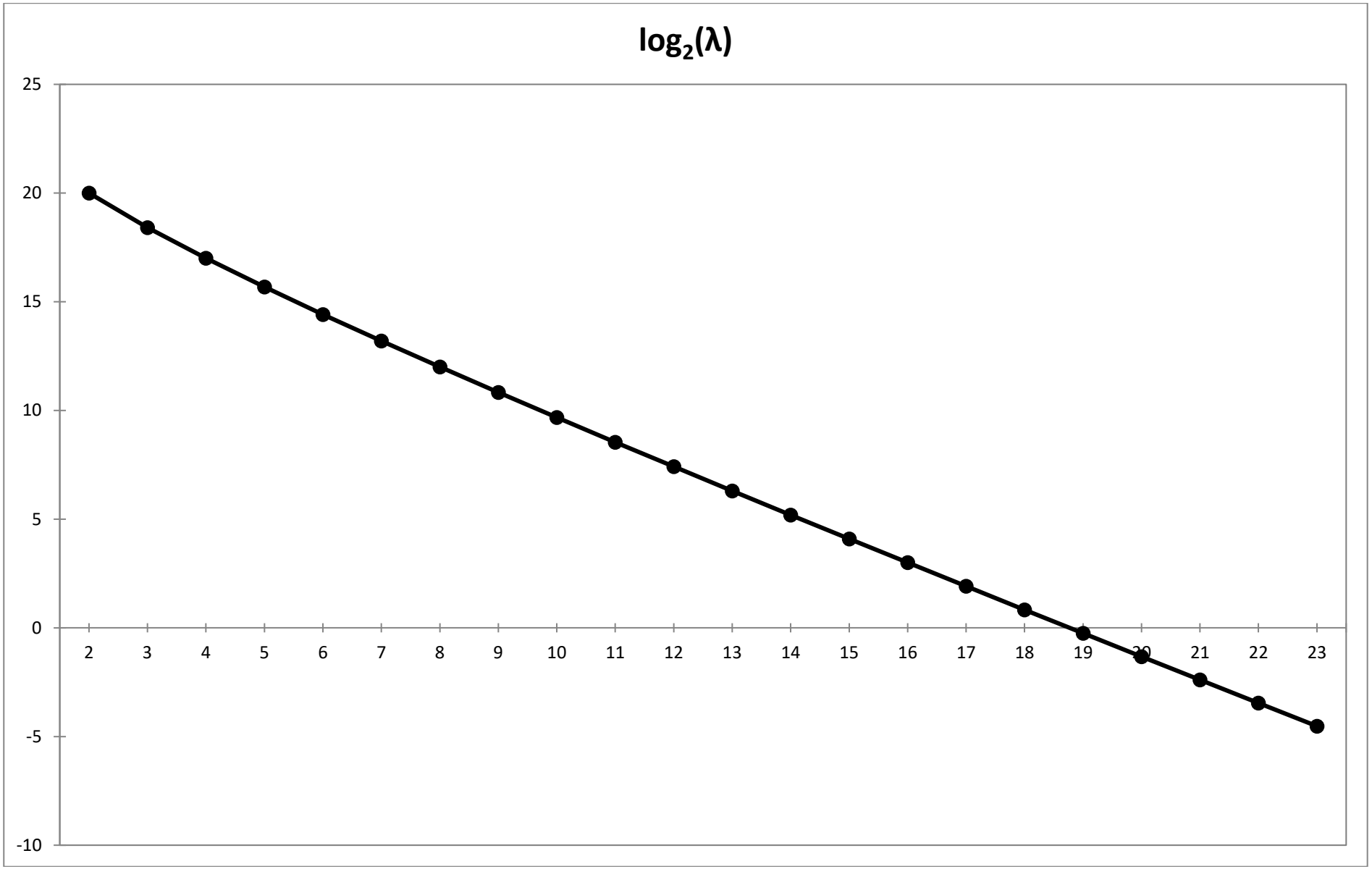

Рисунок 5 – Зависимость λ от *s* в логарифмической шкале

Вторая серия экспериментов проводилась при фиксированном значении $\lambda = 2$ (при каждом значении s совпадал начальный фрагмент рассматриваемой последовательности); результаты экспериментов представлены на рисунках 6–9. Как видно из рисунков, статистические оценки энтропии Тсаллиса генераторов G_1, G_2, G_3 выходят за границы интервала (–1; 1), что означает отклонение гипотезы о том, что наблюдаемая последовательность является РРСП. Возвращение внутрь доверительного интервала при больших значениях порядков s объясняется недостаточной длиной последовательности T. Что касается генератора G_4, то выходы за границы доверительного интервала незначительны и позволяют принять гипотезу H_* при $s \leq 23$.

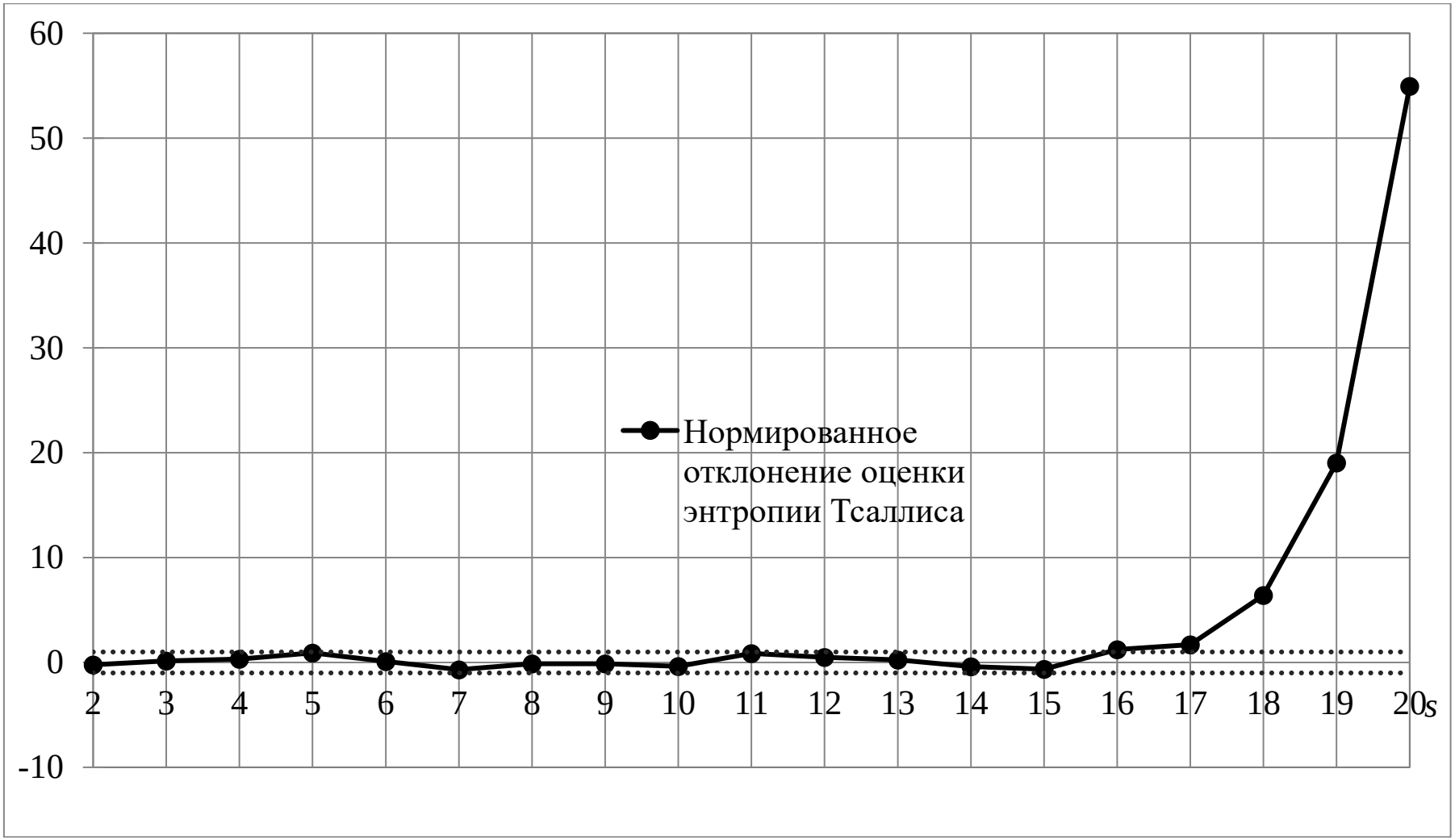

Рисунок 6 – Нормированное отклонение оценки энтропии нелинейного регистра сдвига

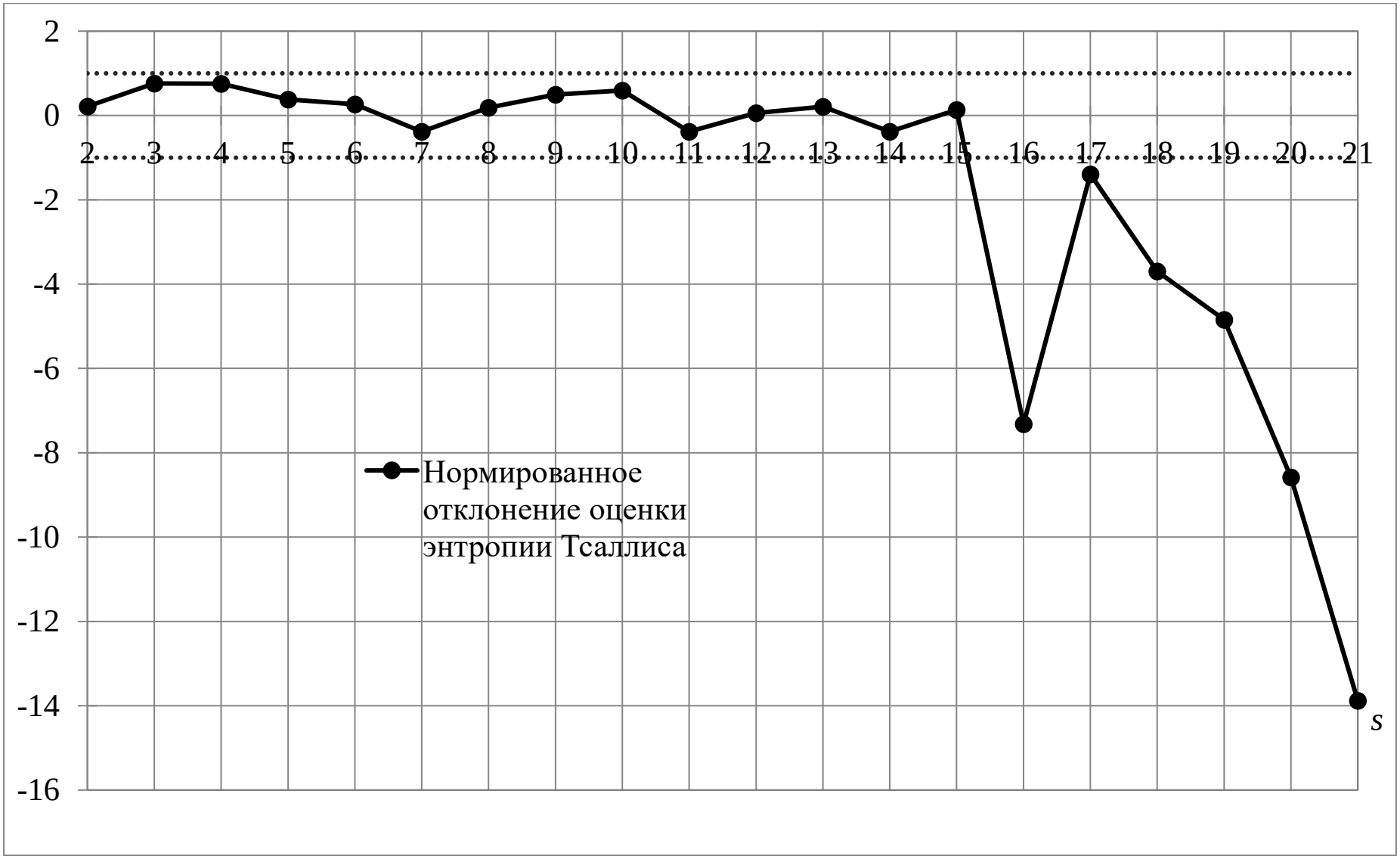

Рисунок 7 – Нормированное отклонение оценки энтропии прореживающего генератора

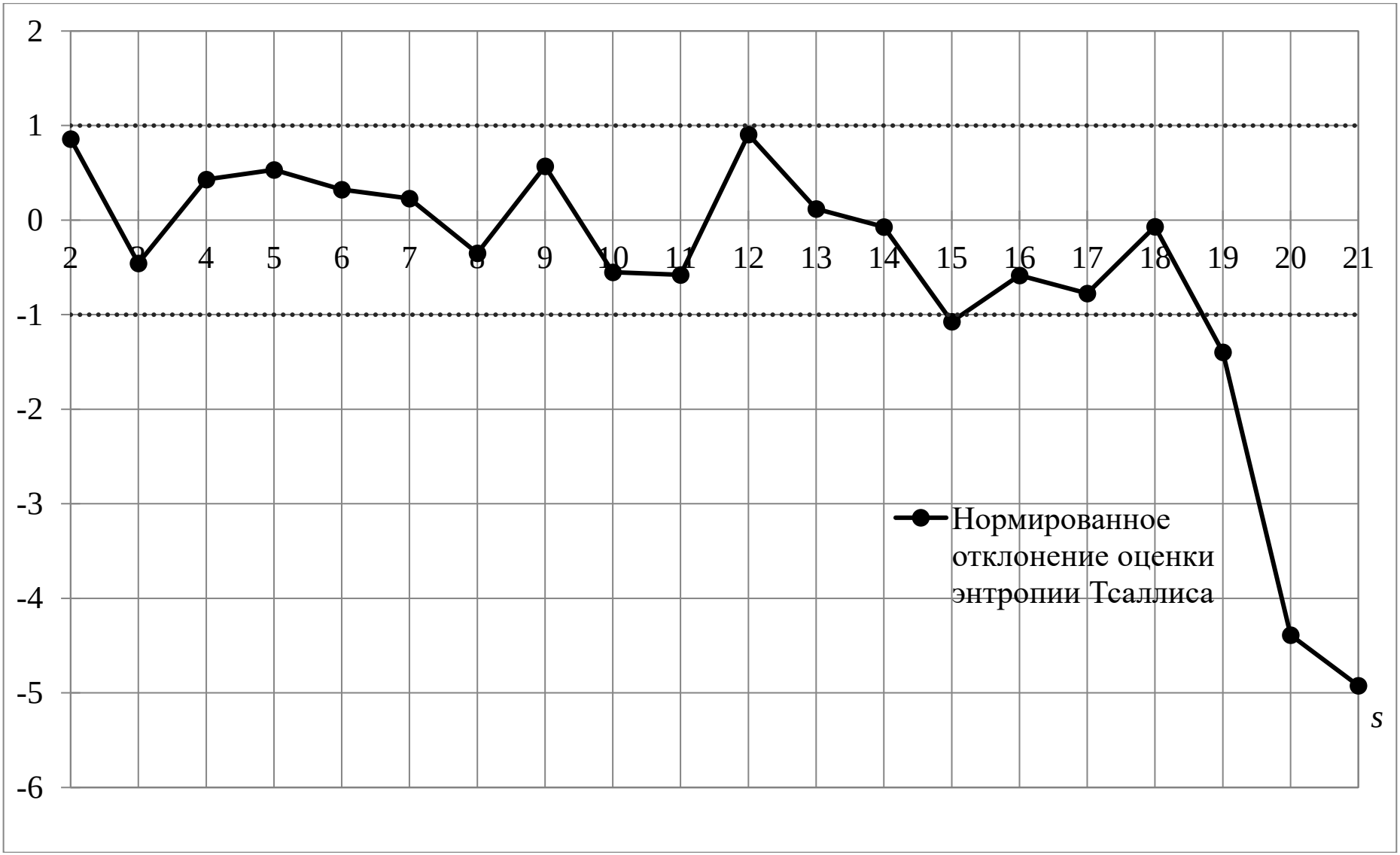

Рисунок 8 – Нормированное отклонение оценки энтропии самосжимающего генератора

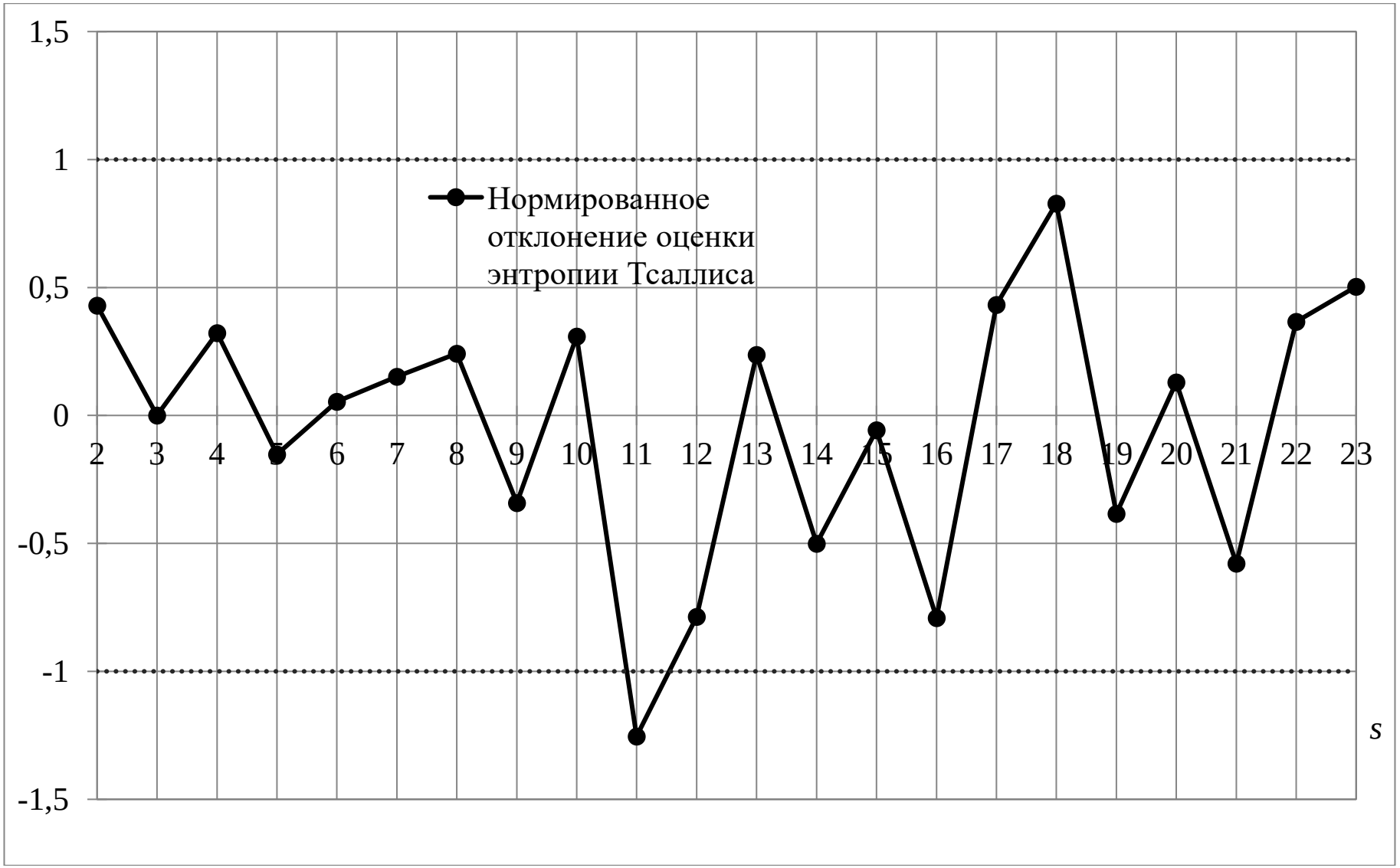

Рисунок 9 – Нормированное отклонение оценки энтропии алгоритма ГОСТ 28147-89

ЗАКЛЮЧЕНИЯ

Разработанный тест на основе статистической оценки энтропии Тсаллиса позволяет обнаруживать отклонения выходных последовательностей псевдослучайных криптографических генераторов от модели РРСП, что обосновывает его применимость для оценки качества псевдослучайных генераторов.

СПАСИБО ЗА ВНИМАНИЕ!