УДК 004.422

ПРИМЕНЕНИЕ РЕСПУБЛИКАНСКОГО РЕГИСТРА «ТУБЕРКУЛЕЗ» В ПРАКТИКЕ РАБОТЫ ФТИЗИАТРИЧЕСКОЙ СЛУЖБЫ

Д. А. Климук✉, Г. Л. Гуревич, Д. М. Журкин, Е. М. Скрягина
Государственное учреждение «Республиканский научно-практический центр пульмонологии и фтизиатрии», Минск, Беларусь
dzklm99@yahoo.com

Введение. Республиканский регистр «Туберкулез» является важным инструментом системы эпиднадзора за туберкулезом в республике. Ведение регистра, согласно приказа Министерства здравоохранения Республики Беларусь, является обязательным для всех противотуберкулезных организаций здравоохранения в республике. Регистр «Туберкулез» является не только средством учета случаев туберкулеза, но также позволяет получать оперативную отчетность в режиме реального времени. В настоящее время регистр содержит сведения о более чем 70 тысячах случаев туберкулеза.

Структура регистра. В своей основе республиканский регистр «Туберкулез» имеет модульную систему – компоненты регистра функционально выделены, однако имеют взаимосвязь.

В 2010 году, согласно приказа Министерства здравоохранения Республики Беларусь, республиканский регистр «Туберкулез» начал функционировать на постоянной основе, представляя на тот момент полицевую базу пациентов с туберкулезом. Впоследствии были разработаны и внедрены дополнительные компоненты регистра.

Компонент по учету пациентов с множественно лекарственно-устойчивым туберкулезом (МЛУ-ТБ) позволил не только вести полицевой учет пациентов с наиболее важной формой туберкулезной патологии, но и предоставил возможности пользоваться сводными таблицами, составленными в соответствии с рекомендуемым набором показателей эпиднадзора за туберкулезом ВОЗ. Компонент лабораторной диагностики позволил накапливать и анализировать данные о статусе бактериовыделения пациентов и тестах лекарственной чувствительности возбудителя, накапливая полную историю пациента, что имеет огромное значение в клинической практике.

Компонент лекарственного менеджмента позволяет получать оперативную информацию об остатках противотуберкулезных лекарственных средств как в республике в целом, так и по определенным противотуберкулезным организациям.

Компонент фармаконадзора позволяет накапливать и анализировать данные о нежелательных лекарственных явлениях, заполняя формы, добавленные в соответствии с нормативно-правовыми документами.

Последний компонент, видеоконтролируемое лечение, является базовым для организации пациент-ориентированного контролируемого амбулаторного лечения туберкулеза.

Перспективы развития. В настоящее время проводится комплекс работ по рефакторингу и реинжинирингу регистра «Туберкулез». Главным нововведением планирует стать компонент по учету назначений схем и лекарственных средств, что позволит получать оперативные данные о применении и учете различных режимов противотуберкулезной терапии. Кроме того, планируется реализовать взаимодействие с другими регистрами в рамках функционирования единой платформы цифрового здравоохранения.

Заключение. Республиканский регистр «Туберкулез» на сегодняшний день представляет собой полноценную медицинскую информационную систему, позволяющую принимать своевременные клинические и управленческие решения. Поддержание и развитие регистра является приоритетным направлением работы противотуберкулезной службы республики.

УДК 004.4: 519.23

ПРОГРАММНОЕ СРЕДСТВО ЭНТРОПИЙНОГО АНАЛИЗА ДИСКРЕТНЫХ ВРЕМЕННЫХ РЯДОВ

В. Ю. Палуха✉, Ю. С. Харин
Учреждение Белорусского государственного университета
«НИИ прикладных проблем математики и информатики», Минск
palukha@bsu.by

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» последовательностью.

Для проверки качества криптографических генераторов используются статистические тесты, в которых проверяется гипотеза H_* = {{x_t} является РРСП}} о том, что наблюдаемая последовательность $\{x_t\}$ является равномерно распределённой случайной последовательностью. В качестве тестовой статистики целесообразно использовать статистические оценки энтропии. Одними из самых распространённых функционалов энтропии являются функционалы Шеннона, Реньи и Тсаллиса, которые и будут рассмотрены в дальнейшем. В НИИ ППМИ разработан программный комплекс, который позволяет вычислять оценки указанных функционалов энтропии дискретной последовательности и на их основе принимать или отклонять гипотезу о «чистой случайности» анализируемой последовательности.

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

В таблице 1 приведены формулы наиболее распространённых функционалов энтропии.

Таблица 1. 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	$S_r(p) = \frac{1}{r-1}\left(1 - \sum_{i=1}^{N} p_i^r \right), r \in \mathbb{N}, r > 1.$

Пусть имеется реализация случайной последовательности $\{x_t : t = 1, \ldots, n\}$ объёма n из распределения вероятностей $\{p_k\}$, по которой будет оцениваться энтропия. Частотные оценки вероятностей имеют вид

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

Рассмотрим асимптотику

$$n, N \to \infty, n/N \to \lambda, 0 < \lambda < \infty. \tag{2}$$

которая отличается от классической ($n \to \infty$, $N < \infty$) тем, что длина последовательности n и мощность алфавита N растут синхронно.

Оценка энтропии Шеннона на основе статистик (1) имеет вид:

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

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

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

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

Из теоремы 1 видно, что в асимптотике (2) оценка (3) является смещённой. Для функционалов энтропии Реньи и Тсаллиса можно построить несмещённую оценку в асимптотике (2), в т.ч. и при $\lambda < 1$.

Определим r-ую нисходящую факториальную степень x:

$$x^{\underline{r}} = x(x-1)\ldots(x-r+1) = \frac{x!}{(x-r)!} = \sum_{i=0}^{r} s(r,i) x^i, \tag{6}$$

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

Статистические оценки энтропии Реньи и Тсаллиса, построенные с использованием нисходящей факториальной степени (6), имеют вид

$$\hat{H}_r(n, N) = \frac{1}{1-r} \ln \left(\sum_{k=1}^{N} \frac{\nu_k^{\underline{r}}}{n^r} \right) = \ln n + \frac{1}{r-1} \left(\ln n - \ln \sum_{k=1}^{N} \nu_k^{\underline{r}} \right), \tag{7}$$

$$\hat{S}_r(n, N) = \frac{1}{r-1} \left(1 - \sum_{k=1}^{N} \frac{\nu_k^{\underline{r}}}{n^r} \right) = \frac{1}{r-1} \left(1 - \frac{1}{n^r} \sum_{k=1}^{N} \nu_k^{\underline{r}} \right). \tag{8}$$

Теорема 2 [3]**.** В асимптотике (2) статистика (8) является состоятельной асимптотически несмещённой оценкой энтропии Тсаллиса и при истинной гипотезе H_* и при $r = 2$ имеет асимптотически нормальное распределение с параметрами:

$$\mu_{S,2} = 1 - \frac{1}{N}, \quad \sigma_{S,2}^2 = \frac{2}{Nn^2}. \tag{9}$$

Теорема 3 [3]**.** В асимптотике (2) статистика (7) является состоятельной оценкой энтропии Реньи и при истинной гипотезе H_* и при $r = 2$ имеет асимптотически нормальное распределение с параметрами:

$$\mu_{H,r} = \ln N, \quad \sigma^2_{H,2} = \frac{2}{n\lambda}. \tag{10}$$

Пусть $\alpha \in (0,1)$ – заданный уровень значимости. Введём обозначения: h – статистическая оценка энтропии Шеннона (3), Реньи (7) или Тсаллиса (8), μ_h и σ_h^2 – соответственно асимптотические математическое ожидание и дисперсия статистической оценки энтропии Шеннона (4), (5), Реньи (10) или Тсаллиса (9) при истинной гипотезе H_* и при $r = 2$. Решающее правило имеет вид [2]:

$$\text{принимается} \ \begin{cases} H_*, & \text{если } t_- < h < t_+; \\ \overline{H_*}, & \text{в противном случае,} \end{cases} \qquad t_\pm = \mu_h \pm \sigma_h \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \tag{11}$$

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

Вычислим нормированную статистику

$$\tilde{h} = \frac{h - \mu_h}{\sigma_h}.$$

Она в асимптотике (2) имеет стандартное нормальное распределение: $\tilde{h} \sim \mathcal{N}(0,1)$. Следовательно, двустороннее *p*-значение для неё равно

$$p - value = 2\left(1 - \Phi\left(\left|\tilde{h}\right|\right)\right). \tag{12}$$

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 (11) мы можем вычислить последовательность нормированных отклонений оценки энтропии от математического ожидания в зависимости от s, которые назовём **энтропийными профилями**:

$$\chi(s) = \frac{\hat{h}(s) - \mu_h(s)}{\sigma_h(s)\Phi^{-1}(1-\alpha/2)} = \frac{\tilde{h}(s)}{\Phi^{-1}(1-\alpha/2)}, \ s = s_-, \ldots, s_+. \tag{13}$$

Программное средство. В НИИ ППМИ разработано программное средство, которое позволяет вычислять оценки энтропии Шеннона (3), Реньи (7) и Тсаллиса (8) при $r = 2$, их асимптотические параметры распределений вероятностей при гипотезе H_* с помощью алгоритмов [4], *p*-значения (12) и энтропийные профили (13) для двоичных файлов. Помимо вывода самих значений, программа выводит графики зависимостей этих величин от длины фрагмента s.

В начале работы необходимо выбрать файл с последовательностью, диапазон $s_-, \ldots, s_+$ и функционалы энтропии. Вычисляемые значения добавляются на экран в режиме реального времени. Имеется возможность изменять уровень значимости

$\alpha \in (0,1)$ без пересчёта оценок энтропии и переключаться на различные режимы отображения: непосредственно оценки энтропии $\hat{h}$, нормированные значения (13), p-значения (12).

Для тестирования программы подготовлена библиотека последовательностей псевдослучайных и физических генераторов. На рисунке 1 представлен результат работы программы с последовательностью физического генератора [5], на рисунке 2 – с последовательностью, полученной при помощи регистра сдвига с линейной обратной связью (РСЛОС) с примитивным характеристическим многочленом над полем GF(2) $x^{32}+x^{22}+x^2+x+1$ [1] на уровне значимости $\alpha = 0.05$. Как видно из рисунков, для физического генератора гипотеза H_* принимается, для РСЛОС начиная с $s = 16$ отклоняется.

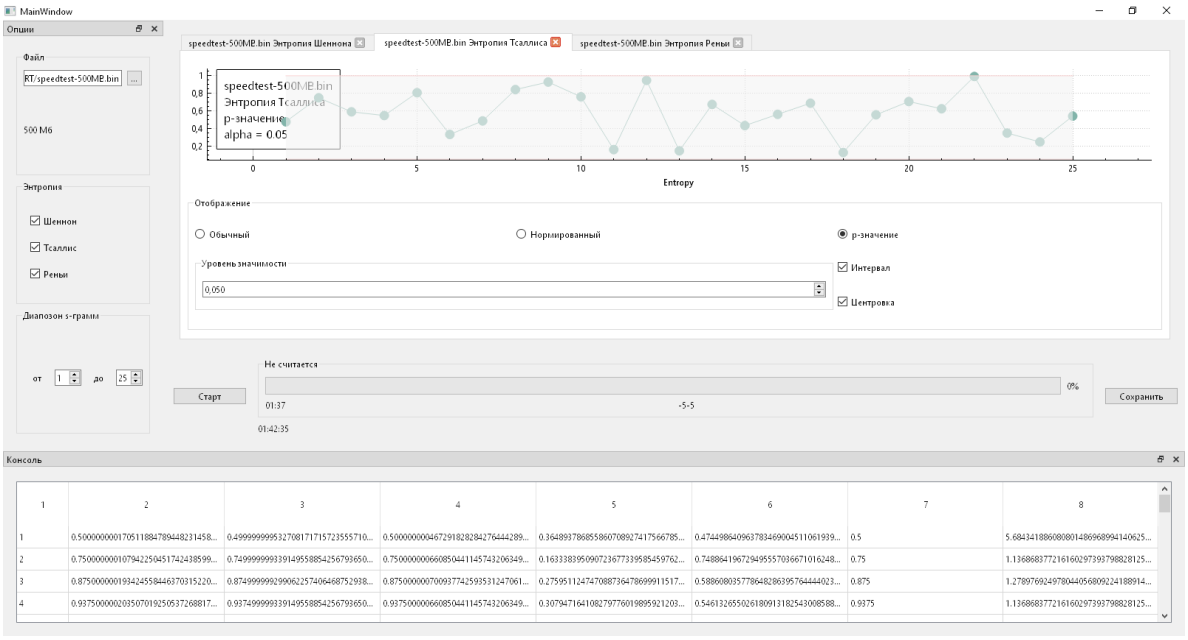

Рис. 1. Энтропийный профиль Тсаллиса физического генератора

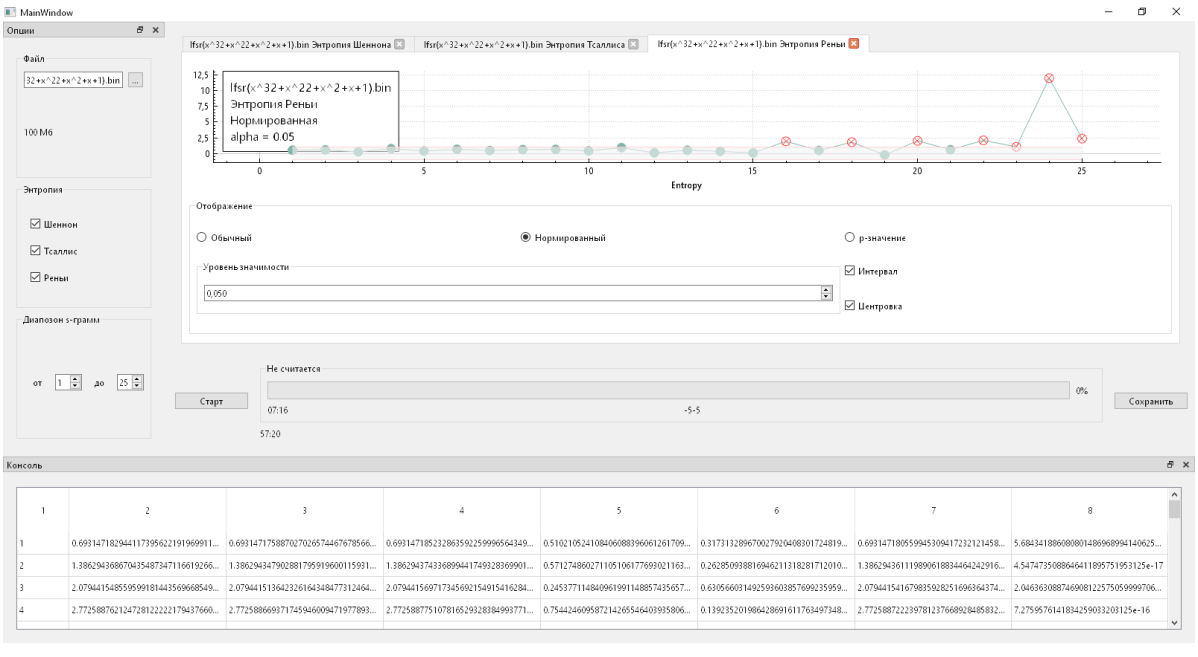

Рис. 2. Энтропийный профиль Реньи РСЛОС

Заключение. На основе исследованного в [2] и [3] энтропийного анализа дискретных временных рядов в НИИ ППМИ разработано программное средство, позволяющее оценивать качество генераторов случайных и псевдослучайных последовательностей на предмет соответствия наблюдаемой последовательности модели «чисто случайной» последовательности. Данная программа является удобной в использовании и позволяет визуализировать полученные результаты.

Список использованных источников

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

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.

3. Харин, Ю. С. 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.

4. Палуха, В. Ю. Вычисление статистических оценок функционалов энтропии двоичных последовательностей / В. Ю. Палуха, Ю. С. Харин // Международный конгресс по информатике: информационные системы и технологии [Электронный ресурс] : Материалы международного научного конгресса. Республика Беларусь, Минск, 24–27 октября 2016 года / редколлегия: С. В. Абламейко (гл. ред.), В. В. Казачёнок (зам. гл. ред.) [и др.]. – Минск: БГУ, 2016. – С. 472–476.

5. 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.