

$J_1^s = (j_1, \dots, j_s) \in V_s$ – s -мерное распределение вероятностей. Построим оценки вероятностей $\{\widehat{p_J}\}$ и их r -ых степеней $\{\widehat{p_J^r}\}$, используя частотные статистики:

$$\widehat{p_J^r}(s) = \frac{v_J^r}{n^r}, \quad v_J = \sum_{k=1}^n \delta_{X^{(k)}, J}, \quad \delta_{X^{(k)}, J} = \begin{cases} 1, & X^{(k)} = J; \\ 0, & X^{(k)} \neq J, \end{cases}$$

и на их основе построим статистические оценки энтропии Шеннона $\widehat{H}(n, s) = - \sum_{J \in V_s} \frac{v_J}{n} \ln \frac{v_J}{n}$,

Реньи $\widehat{H}_r(n, s) = \frac{1}{1-r} \ln \left(\sum_{J \in V_s} \widehat{p_J^r}(s) \right)$ и Тсаллиса $\widehat{S}_r(n, s) = \frac{1}{r-1} \left(1 - \sum_{J \in V_s} \widehat{p_J^r}(s) \right)$, $r \in \mathbb{N}$, $r > 1$, которые в асимптотике $n, N = 2^s \rightarrow \infty, n/N \rightarrow \lambda, 0 < \lambda < \infty$ имеют асимптотически нормальное распределение при гипотезе H_* [1].

Зададим уровень значимости $\alpha \in (0, 1)$. Введём обозначения: $\widehat{h}(s)$ – статистическая оценка энтропии Шеннона, Реньи или Тсаллиса, $\mu_h(s)$ – асимптотическое математическое ожидание, $\sigma_h^2(s)$ – асимптотическая дисперсия оценки при истинной гипотезе H_* , $\Phi^{-1}(\cdot)$ – квантиль стандартного нормального закона. Вычислим для наблюдаемой последовательности статистику $\widehat{h}(s)$. Решающее правило, основанное на статистике $\widehat{h}(s)$, имеет вид:

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

На основе решающего правила (1) мы можем вычислить последовательность нормированных отклонений оценки энтропии от математического ожидания в зависимости от s , т.е. последовательность величин

$$\chi(s) = \frac{\widehat{h}(s) - \mu_h}{\sigma_h(s) \Phi^{-1}(1 - \alpha/2)}, \quad s = 1, \dots, s_+,$$

которую будем называть энтропийным профилем. Тестирование с помощью профиля позволяет выносить решение о принятии или отклонении гипотезы H_* на основе решающего правила (1) по последовательности значений $\chi(s)$ для различных s ; такое решение видится более аргументированным, чем при принятии его по результатам однократного применения теста (1) для одного значения s .

Литература

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

ТЕСТИРОВАНИЕ ИСТОЧНИКОВ СЛУЧАЙНОСТИ НА ОСНОВЕ ОЦЕНОК МИНИМАЛЬНОЙ ЭНТРОПИИ

Пирштук И.К.

Белгосуниверситет, НИИ прикладных проблем математики и информатики
Независимости 4, 220030 Минск, Беларусь, pirshuk@bsu.by

Введение. Постановка задачи. Случайные числа (СЧ) широко применяются в криптографических приложениях для формирования криптографических ключей, синхросылок и других объектов с целью передачи зашифрованных данных. СЧ могут быть получены различными способами, чаще всего для генерации СЧ используются генераторы случайных чисел (ГСЧ). Рекомендуются использовать физические ГСЧ (ФСЧП). Требования к выходным

последовательностям ГСЧ достаточно жесткие, а природа данных может быть неизвестна. Для проверки статистических качеств выходных последовательностей ГСЧ используются специальные наборы статистических тестов. Наиболее полное описание таких наборов приведено в пакете статистических тестов Национального института стандартов и технологий США (NIST). Эти тесты основаны на различных статистических свойствах, присущих только равномерно распределенным случайным последовательностям.

При оценке качеств ГСЧ, наряду с оцениванием случайности, актуальной является задача оценивания энтропии выходных данных источника случайности (ядра ГСЧ). Для вычисления энтропии рекомендуется выбирать заниженную меру — минимальную энтропию [1]. При таком подходе источник случайности ГСЧ гарантирует заданный уровень энтропии.

Статистические тесты оценивания минимальной энтропии включают 10 тестов [1]: 1) коллизий, 2) неполных коллекций, 3) Маркова, 4) сжатия, 5) частотного, 6) наиболее повторяющейся подстроки, 7) множественного прогнозирования наиболее частотного значения в окне, 8) прогнозирования задержки, 9) множественного прогнозирования моделей Маркова и 10) прогнозирования *LZ78Y*. Все тесты реализованы нами на ПЭВМ, протестированы на модельных данных и применены для реальных данных ФГСЧ и других источников случайности.

Результаты численных экспериментов включают оценки минимальной энтропии случайных данных различной природы. В качестве иллюстрации приведены результаты обработки 5 бинарных выборок длительностью 1 Мб каждая (2^{23} бит), полученных с ФГСЧ на основе шумового диода. Для каждой выборки были применены по 5 лучших тестов (в смысле быстродействия). Результаты оценивания приведены ниже в таблице. Легко видеть, что в данном случае минимальная энтропия равна 0.938882.

Номер	частотный	коллизий	Маркова	сжатия	прогн.задержки
1	0.998235	0.969419	0.998718	0.986677	0.998165
2	0.998607	0.944887	0.999118	0.967438	0.998760
3	0.998554	0.938882	0.998030	0,953643	0.998930
4	0.998088	0.951598	0.998615	0,971231	0.999205
5	0.998370	0.965373	0.999000	0,970713	0.998886

Выводы и рекомендации по применению тестов включают: 1) определение минимального размера выборки для каждого теста; 2) исследование быстродействия тестов в зависимости от мощности алфавита и размера выборки; 3) условие вычисления гарантированного значения оценки минимальной энтропии (следует использовать не менее 5 тестов).

Литература

1. NIST Special Publication 800-90B (SP 800-90B). Recommendation for the Entropy Sources Used for Random Bit Generation. — January 2018.